Volume 65D

Number 3

May–June 1961

Radio Propagation

SECTION D

JOURNAL OF RESEARCH

NATIONAL BUREAU OF STANDARDS

Journal of Research

of the

National Bureau of Standards

D. RADIO PROPAGATION

MAY–JUNE • 1961

VOLUME • 65D

NUMBER • 3

Publication dates:
Jan. 15, Mar. 15, May 15, July 15, Sept. 15 ,Nov. 15, 1961

JOURNAL OF RESEARCH

The National Bureau of Standards Journal of Research reports research and development in the fields of activity shown at right. Also included from time to time are survey articles related to the Bureau's scientific and technical program.

The Journal is published in four separate sections as follows:

A. Physics and Chemistry

Contains papers of interest primarily to scientists working in these fields. Issued six times a year. Annual subscription: domestic, $4.00; foreign, $4.75.

B. Mathematics and Mathematical Physics

Presents studies and compilations designed mainly for the mathematician and the theoretical physicist. Issued quarterly. Annual subscription: domestic, $2.25; foreign, $2.75.

C. Engineering and Instrumentation

Reports research and development results of interest chiefly to the engineer and the applied scientist. Issued quarterly. Annual subscription: domestic, $2.25; foreign, $2.75.

D. Radio Propagation

Reports research in radio propagation, upper atmospheric physics, and communications. Issued six times a year. Annual subscription: domestic, $4.00; foreign, $4.75.

Functions and Activities

The functions of the National Bureau of Standards are set forth in the Act of Congress, March 3, 1901, as amended. These include the development and maintenance of the national standards of measurement and the provision of means and methods for making measurements consistent with these standards; the determination of physical constants and properties of materials; the development of methods and instruments for testing materials, devices, and structures; advisory services to government agencies on scientific and technical problems; invention and development of devices to serve special needs of the Government; and the development of standard practices, codes, and specifications. The work includes basic and applied research, development, engineering, instrumentation, testing, evaluation, calibration services, and various consultation and information services. The Bureau also serves as the Federal technical research center in a number of specialized fields. The scope of activities of the National Bureau of Standards is suggested in the following listing of the divisions and sections engaged in technical work.

Washington, D.C.

Electricity. Resistance and Reactance. Electrochemistry. Electrical Instruments. Magnetic Measurements. Dielectrics.

Metrology. Photometry and Colorimetry. Refractometry. Photographic Research. Length. Engineering Metrology. Mass and Scale. Volumetry and Densimetry.

Heat. Temperature Physics. Heat Measurement. Cryogenic Physics. Equation of State. Statistical Physics.

Radiation Physics. X-Ray. Radioactivity. Radiation Theory. High Energy Radiation. Radiological Equipment. Nucleonic Instrumentation. Neutron Physics.

Analytical and Inorganic Chemistry. Pure Substances. Spectrochemistry. Solution Chemistry. Analytical Chemistry. Inorganic Chemistry.

Mechanics. Sound. Pressure and Vacuum. Fluid Mechanics. Engineering Mechanics. Rheology. Combustion Controls.

Organic and Fibrous Materials. Rubber. Textiles. Paper. Leather. Testing and Specifications. Polymer Structure. Plastics. Dental Research.

Metallurgy. Thermal Metallurgy. Chemical Metallurgy. Mechanical Metallurgy. Corrosion. Metal Physics.

Mineral Products. Engineering Ceramics. Glass. Refractories. Enameled Metals. Crystal Growth. Physical Properties. Constitution and Microstructure.

Building Research. Structural Engineering. Fire Research. Mechanical Systems. Organic Building Materials. Codes and Safety Standards. Heat Transfer. Inorganic Building Materials.

Applied Mathematics. Numerical Analysis. Computation. Statistical Engineering. Mathematical Physics.

Data Processing Systems. Components and Techniques. Digital Circuitry. Digital Systems. Analog Systems. Applications Engineering.

Atomic Physics. Spectroscopy. Radiometry. Solid State Physics. Electron Physics. Atomic Physics.

Instrumentation. Engineering Electronics. Electron Devices. Electronic Instrumentation. Mechanical Instruments. Basic Instrumentation.

Physical Chemistry. Thermochemistry. Surface Chemistry, Organic Chemistry. Molecular Spectroscopy. Molecular Kinetics. Mass Spectrometry. Molecular Structure and Radiation Chemistry.

• Office of Weights and Measures

Boulder, Colo.

Cryogenic Engineering. Cryogenic Equipment. Cryogenic Processes. Properties of Materials. Gas Liquefaction.

Ionosphere Research and Propagation. Low Frequency and Very Low Frequency Research. Ionosphere Research. Prediction Services. Sun-Earth Relationships. Field Engineering. Radio Warning Services.

Radio Propagation Engineering. Data Reduction Instrumentation. Radio Noise. Tropospheric Measurements. Tropospheric Analysis. Propagation-Terrain Effects. Radio-Meteorology. Lower Atmospheric Physics.

Radio Standards. High-Frequency Electrical Standards. Radio Broadcast Service. Radio and Microwave Materials. Atomic Frequency and Time Interval Standards. Electronic Calibration Center. Millimeter-Wave Research. Microwave Circuit Standards.

Radio Systems. High Frequency and Very High Frequency Research. Modulation Research. Antenna Research. Navigation Systems. Space Telecommunications.

Upper Atmosphere and Space Physics. Upper Atmosphere and Plasma Physics. Ionosphere and Exosphere Scatter. Airglow and Aurora. Ionospheric Radio Astronomy.

U.S. DEPARTMENT OF COMMERCE
Luther H. Hodges, Secretary

NATIONAL BUREAU OF STANDARDS
A. V. Astin, Director

Order all publications from the Superintendent of Documents, U.S. Government Printing Office, Washington 25, D.C.

Use of funds for printing this publication approved by the Director of the Bureau of the Budget (June 24, 1958).

Editorial Notice

Conference on Transmission Problems Related to High-Frequency Direction Finding

This issue includes a group of papers presented in June 1960 at a conference sponsored by the University of California at Los Angeles in cooperation with the Office of Naval Research. The purpose of the conference was to discuss the aspects of long-range high-frequency radio propagation that affect radio location and direction finding, and the related problems of measurement and analysis.

A comprehensive bibliography of published work on direction finding and related ionospheric propagation topics for the period 1955–1959 has been prepared by the Numerical Analysis Research Staff of the University of California at Los Angeles. This bibliography will be edited by the Radio Systems Division of the NBS and published as a Technical Note of the NBS.

THOMAS N. GAUTIER, *Associate Editor.*

JOURNAL OF RESEARCH of the National Bureau of Standards—D. Radio Propagation
Vol. 65D, No. 3, May–June 1961

Propagation Studies Using Direction-Finding Techniques[1,2]

Edgar C. Hayden

(August 26, 1960)

The most persistent and difficult problem in radio direction finding has been measurement of the direction of arrival of an incident signal field under circumstances where multipath propagation is possible. In the HF band, the ionosphere plays a predominant role in the propagation of radio waves, and in this region several mechanisms exist which promote splitting of a radio signal into numerous components.

It would be of value to have a more thorough knowledge of the characteristics of the individual signal components. In this paper two techniques are described for the study of multicomponent signals. One involves use of pulse transmissions to effect "time-of-arrival" resolution; the other involves use of a highly directive antenna system to effect "direction-of-arrival" resolution.

Results of the application of these two techniques in specific instances are presented.

1. Introduction

In the HF band (3 to 30 Mc/s) long-distance transmission is possible with moderate power levels because of the existence of the ionosphere. However, while the ionosphere makes such transmission possible, it at the same time has characteristics which make difficult the measurement of signal direction of arrival and the extrapolation of such measurements for position fixing or propagation studies. The chief of these difficulties is that frequently there are in fact several possible ray paths between transmitter and receiver, and these paths do not lie exactly in the plane of the great circle passing through the transmitter and receiver locations. The former phenomenon complicates the direction-of-arrival measurement problem, and the latter the problem of extrapolation of the direction-of-arrival measurements to the point of origin of the signal.

The mechanisms giving rise to multiple paths through the ionosphere; i.e., the stratification of the region, magneto-ionic splitting, and irregularities of structure, are well known in principle, if not in details of behavior.

As a consequence of the multiple signal rays arriving at a receiving point, the field in the neighborhood of the receiving antenna exhibits "interference" or "fringe" effects. Most direction-measuring techniques give an indication which approximates the normal to the equiphase surface in the immediate neighborhood of the antenna system. In such an interference field, the equiphase surfaces are corrugated rather than plane, so that the indicated direction of arrival depends on the location of the instrument in the field. Since the field pattern moves about as the relative phases of the various rays change, the indication will fluctuate. Thus, for multiray fields, the normal to the equiphase surface is not a satisfactory definition of "direction of propagation" for purposes of ray retracing.

The nature of the problem is illustrated in figure 1.

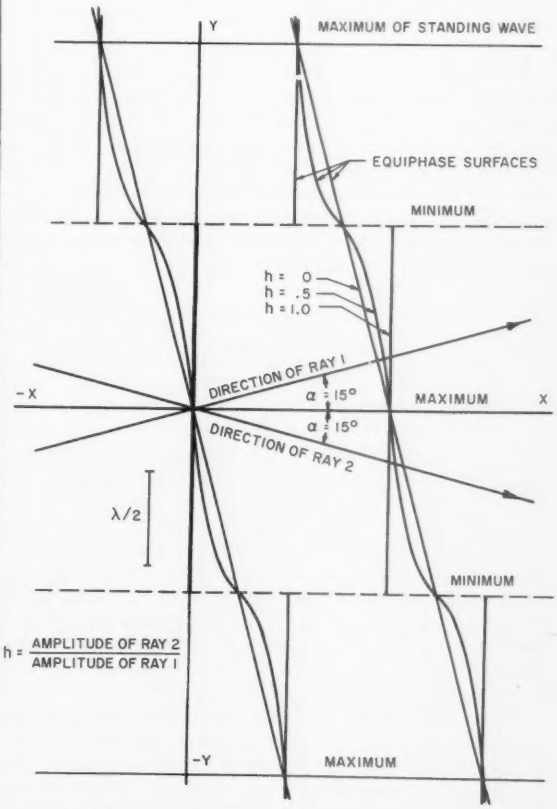


FIGURE 1. *Wave interference pattern for ray separation of 30°.*

[1] Contribution from Radio Direction-Finding Research Group, Department of Electrical Engineering, University of Illinois, Urbana, Ill.

[2] Paper presented at the Conference on Transmission Problems Related to High-Frequency Direction Finding, at UCLA, June 21–24, 1960.

When two waves, each of which is assumed to be a plane wave with its wave normal lying in the plane of the paper, arrive at a receiving site from different directions, a standing wave field is produced. If (as on a transmission line) the two are going in opposite directions, the maxima and minima are each spaced one-half wavelength, and lie in planes perpendicular to the directions of propagation.

The equiphase surfaces are also planes perpendicular to the direction of propagation. If the angle between the directions of propagation of the waves is reduced from 180° toward zero (it is 30° in figure 1), the maxima and minima remain plane surfaces which are parallel to the bisector of the angle between the wave normals or rays. Their spacing, however, increases. It is

$$s_{M-M}=s_{m-m}=\frac{\lambda}{2\sin\alpha}$$

s_{M-M} is the distance between maxima

s_{m-m} is the distance between minima

λ is the wavelength

2α is the angle between the wave normals or rays (1)

Unlike the maxima and minima, the equiphase surfaces do not remain plane. They become corrugated (or, if the two waves are of exactly equal strength, broken). "On the average," they are perpendicular to the wave normal of the stronger ray. If the waves are assumed to be in phase at the origin, the intersections of the equiphase surfaces with the XY plane (the plane of the paper) are given by the following expression:

$$Y=-\frac{\lambda}{2\pi\sin\alpha}\arctan\left[\frac{1+h}{1-h}\tan\frac{(2\pi X\cos\alpha)}{\lambda}\right]-\frac{n\lambda}{\sin\alpha}$$

$$n=\ldots -3,-2,-1,0,+1,+2,\ldots \qquad (2)$$

It can be seen in figure 1 that the normal to the phase front varies from a direction between the wave normals (but closer to the normal to the stronger wave), to a direction outside the angle between the wave normals (but away from the normal to the weaker wave). Thus, as the relative phase between the waves changes, the normal to the equiphase surface at a fixed point will swing back and forth, across the direction of the normal to the stronger wave. As the two signals approach equality in strength, the swing approaches 90°.

When the angle between the wave normals (2α) is fairly small, the distance between the corrugations of the phase surfaces is several times as great as the dimensions of a small aperture system. For example, if α is 5°, the distance between adjacent corrugations is over 5.5λ. This is about 22 times the maximum useful size [3] of a 4-element Adcock system, and about 5.5 times the maximum useful size [3] of a regularly spaced 8-element Adcock system. All present small aperture direction-finding systems give an indicated bearing which is equal to, or very nearly equal to, the normal to the equiphase surface (this is the commonly used definition of the direction of propagation). Thus these small systems can give directional indications which are considerably (up to 90° in the extreme case) in error. The indicated bearing can show considerable fluctuation even though the actual ray arrival angles are steady.

In figure 2 the error in indicated bearing is shown graphically for the particular case in which 2α is 30°. Curves are given for several values of the relative amplitude of the waves. The error curves for other values of α are similar in character. Note that as the relative phase varies, the deviation to one side of the stronger ray is less than that to the other side.

Similar effects occur in systems using arrays of large aperture, though the error fluctuations can in general be confined to a smaller range. They may, however, be considerably greater than the true fluctuations in direction of arrival of the rays.

Thus, under ordinary circumstances, direction-finding systems of both large and small aperture are subject to errors in bearing indication when the incident signal is multirayed. If, then, the characteristics of an individual ray are to be studied, some means must be provided for ray separation; i.e., for taking the signal apart into its components. Three possibilities are immediately apparent:

(1) Wait until only one ray is present before acquiring data.
(2) Take advantage of differences in ray path length to effect a resolution on the basis of time of arrival.
(3) Take advantage of differences in ray path direction to effect a separation on the basis of direction of arrival.

The first of these very severely restricts the situations under which data can be acquired. In practice it is of limited utility. The second requires the use of signals consisting of short pulses, the resolution capability being determined by the pulse duration. The third requires large antenna systems of high directivity. The resolution capability is limited by the beamwidth of the array radiation pattern.

The latter two of the techniques listed are both useful. Each covers some situations in which the other fails. In particular, the time-of-arrival technique is often satisfactory for resolution of the normal modes of propagation, but fails under anomalous conditions because of the great multiplicity of raylets and the extremely short pulse duration which would be required for resolution. On the other hand, the

[3] The maximum useful size of these systems is determined by the maximum permissible spacing error. The figures above are for systems whose maximum spacing error is limited to about 2°.

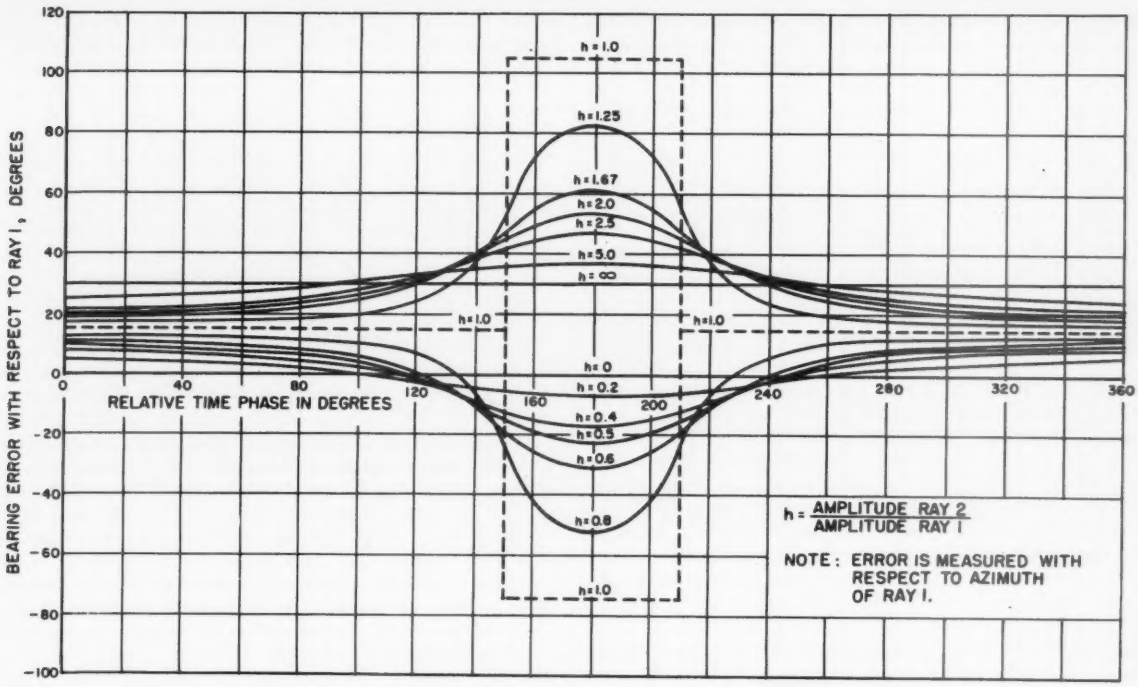


FIGURE 2. *Wave interference bearing error for two-ray signal with ray separation of 30°.*

direction-of-arrival technique can often give useful information under anomalous conditions because the angular spectrum of such signals is frequently of large extent, but may fail under normal conditions because the angular separation of the rays is small. Some of the advantages of each technique have been exploited. The results of two specific experiments are reported in the following sections.

2. Ray Resolution by Time-of-Arrival

The technique used in this experiment for taking apart a multiray signal involves exploitation of the differences in time of arrival of the various rays. Such a technique, while not suited to the general run of signals, is well-suited for use under controlled circumstances. With a cooperative transmitter, the signal modulation characteristic can be chosen to facilitate the receiving problem.

For resolution of multiray signals a short pulse is required. Since direction-finding techniques must often be used with transmissions which are modulated by waveforms other than pulses, it would be useful to provide a signal to simulate other sorts of transmissions as well. Comparisons would thus be facilitated between measurements of the directional characteristics of the component modes, and the directional indications given on the composite signal by various instruments. Such a signal can be provided by transmitting alternately a pulse short enough to resolve the component modes and a pulse long enough that the received pulses for the various modes overlap an appropriate amount.

Choice of a length for the short pulse is a compromise between the resolution required on one hand, and signal-to-noise ratio, available receiver bandwidth, Federal Communications Commission regulations, and the overcrowded state of the HF band on the other hand. The latter two considerations seemed to be the deciding factors. In order to avoid interference from other transmitters on adjacent channels, and to avoid interference with other users of the spectrum, the bandwidth was limited to 25 kc/s. For double-sideband transmission the minimum usable pulse width under these conditions is about 80 μsec. Fortunately this was short enough, for the transmitter location chosen, to resolve the grossly different modes, though it was only occasionally that the two magneto-ionic components of a given mode could be resolved completely.

The greatest expected difference in time of arrival between modes of any consequence was of the order of 4 msec. A long pulse of 12 msec duration was considered adequate to insure overlap of all interesting modes, and to provide a composite signal of sufficient duration.

The modulation waveform chosen for the transmitted signal is shown in figure 3a. If a signal of this configuration were to travel to a receiving location via two paths, the structure of the received signal

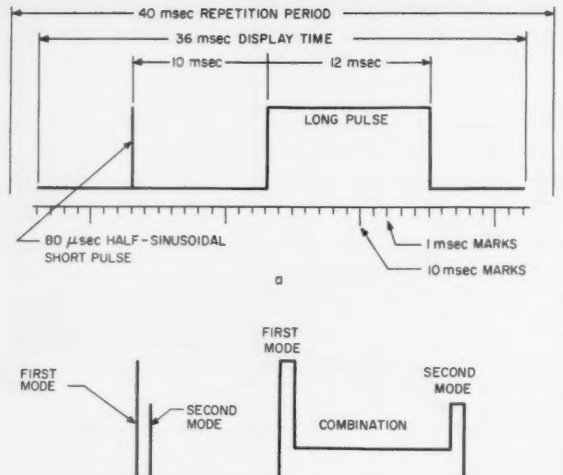


FIGURE 3. *Signal envelope waveforms.*

a. Transmitted waveform.

b. Example of a received waveform; two modes, amplitude of first 3/2 that of second, relative phase 180°, difference in arrival time 1 msec.

c. Expanded sweep, short pulses only.

might, for example, be as shown in figure 3b. The long pulse shows three regions, a leading edge in which only the first-to-arrive ray is present, a trailing edge in which only the last-to-arrive ray is present, and a central region in which a composite of the rays is present. Thus the long pulse permits some decomposition of the signal, but also permits study of the composite signal. If the difference in time of flight for the two rays is more than the duration of the short pulse, then two distinct short pulses will be observed (fig. 3c). In some cases each of these short pulses can be further split into magneto-ionic components, thus allowing an additional step in the decomposition of the signal.

Because of its ability to work on pulse signals, the directional measurement system chosen was a conventional Adcock antenna system followed by a twin-channel receiver and a cathode-ray-tube goniometer. The elements of the system are shown in figure 4.

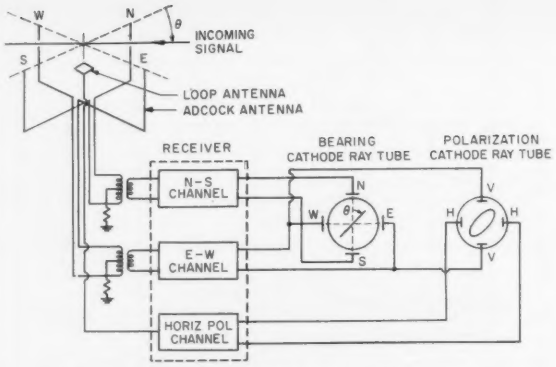


FIGURE 4. *Direction-finding system, block diagram.*

In addition to the bearing, information about the state of polarization was needed. A horizontal loop antenna was installed in the center of the antenna array. It can be seen in figure 4. Its output was received by a third channel of the direction-finding receiver. Since the direction-finding antenna was vertically polarized, and since signals were expected to come from an easterly direction, the output of the *E–W* channel of the direction-finding receiver was used, with the output from the loop receiver, to display the state of polarization on a cathode ray tube. Quantitative measurement of the state of polarization was not needed, so this simple system sufficed in spite of the different directional characteristics of the horizontally and vertically polarized parts of the system.

Three cathode ray tubes, as shown diagrammatically in figure 5, were used in the display system, one to display bearing, a second polarization, and the third signal structure. The third tube was a two-gun tube, one gun being used to display the vertically polarized signal, and the other the horizontally polarized signal. In addition, each gun was double-traced, alternate traces serving to display signal structure and timing scale. A mode selection gating pulse, variable in

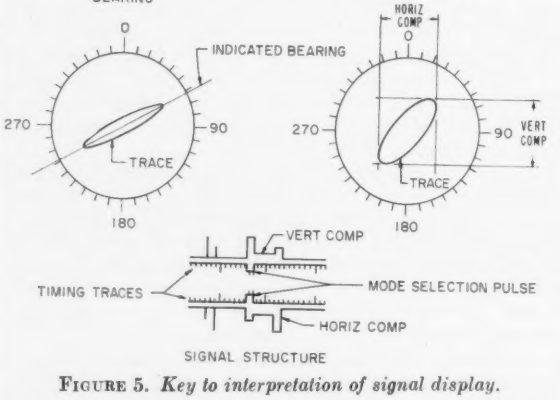


FIGURE 5. *Key to interpretation of signal display.*

duration and in position along the structure trace, was provided to select portions of the signal for display on the bearing and polarization indicators. The duration and position of this gate are shown as a "step" in the timing scale base line. In some cases an expanded structure sweep was used, showing only the short pulses as illustrated in figure 3c. The expansion factor is roughly 10 X.

Transmitter pulse repetition rate and receiver display rate were synchronized by deriving each of them from a stable 100 kc/s oscillator. The basic pulse repetition rate was 25/sec.

The displayed information was recorded with a specially constructed pulse-driven 16-mm camera, whose operation was controlled by the timing pulse train. The next few figures show examples of typical signal displays. The figures were made from individual frames of the actual 16-mm film recording. Reference to figures 3 and 5 will aid in understanding figures 6, 7, 8, and 9. In the figures showing the complete signal waveform, the camera shutter was open for one data cycle (2/25 sec because of the double-tracing of the structure display to present both timing and data traces). The shutter was opened and closed in the sweep flyback periods. Thus one timing and signal structure trace, but two successive bearing and polarization traces, were recorded. In the figures showing the short pulses only, the shutter was open for five complete data cycles (10/25 sec), causing five timing and signal structure traces, but ten bearing and polarization traces, to be recorded. The multiple exposure of the short pulse signals was of material help in reducing the effects of noise on the bearing display.

Figure 6 shows a two-mode signal (one-hop E and one-hop F). The two short pulses are visible near the left end of the structure display traces. The long pulse shows three portions, a portion where only the first mode is present, a portion of composite signal, and a portion where only the second mode is present. The mode selection pulse is set to select the first mode and a portion of the composite signal. Bearings and polarization ellipses for these two portions of the signal appear superposed on the appropriate indicators. The two superposed images on each of the indicators may be differentiated by comparison of their relative dimensions with the relative amplitudes of the signal components as shown on the structure display. Note that only the vertically polarized (upper) component of the structure display is of use in interpreting the bearing display, since the direction-finding antenna was vertically polarized. Both components of the structure display are useful in interpreting the polarization display. This series of five pictures shows one fading cycle, as the relative phase of the components takes on successive values of approximately 0°, 90°, 180°, 270°, and 360°. The steadiness of the bearing and polarization indications on the single mode should be contrasted with the variability of these indications on the composite portion of the signal. The behavior of the bearing indication on the composite portion of the signal should be compared with the wave interference bearing error curves in figure 2.

Figure 7 shows a similar fading cycle, but the mode selection pulse is set to pick out the second mode, rather than the first, and a portion of the composite signal.

Figures 8 and 9 show a one-hop F-layer signal. The structure trace is in the "fast sweep" position so that only the short pulses are displayed.

In figure 8a only the ordinary magneto-ionic component is present, while in 8b only the extraordinary component is present. These two can be differentiated by their characteristic polarizations, which are quite stable from day to day. Note the difference in indicated bearing. In figure 8c both components are present, and are nearly resolved. Both characteristic polarizations are evident. The bearing display is not easily interpretable because the bearings of the two components are so nearly alike that the indications are not distinct.

Figure 9a shows a case in which both magneto-ionic components are present, but are incompletely resolved. In the polarization display the figure starts to build up in one of the characteristic polarizations, then switches to the other. During the transition, or overlap, the vertically polarized components are in phase, but the horizontally polarized components are out of phase. This causes the vertical hourglass shape of the polarization display and propeller shape of the bearing display. This also shows up in the splitting of the horizontally polarized pulse, but not of the vertically polarized pulse. The bearing trace is in effect, swinging from the bearing of one component to that of the other through a small angle. In figure 9b, the situation is similar, except that the vertical components are out of phase and the horizontal components are in phase. In this case the hourglass figure on the polarization display is horizontal. The vertically polarized pulse is split, but the horizontally polarized one is not. The bearing trace swings from the bearing of one component to that of the other through an angle of nearly 180°, giving the effect of a split bearing. The wave interference bearing error curves (fig. 2) will assist in interpreting the bearing displays in figures 9a and 9b. Note in figure 2 that if two components are in phase, and their relative amplitude swings from zero to infinity, the bearing will swing from that for one signal to that for the other through a small angle, while if the components are out of phase, the bearing will swing through the supplement of this angle. It is this action that is taking place in figures 9a and 9b. Figure 9c shows a case where the overlap is so nearly complete that neither the horizontally nor the vertically polarized pulses, are split. However, the presence of the two components can still be detected by the X-shaped figure in the center of the polarization display (compare with figs. 8a and 8b in which only one component was present).

The results reported here are for a particular single path, and a particular frequency. Though

they are for this reason somewhat specialized, they do show some of the phenomena that are observable, and do demonstrate the power of the technique. The transmitter was located at Columbus, Ohio, and the receiver at Urbana, Illinois. Both stations lie within a few minutes of the 40th parallel of latitude. They are separated by a distance of 450 km. The frequency used was 5155 kc/s. The transmitter delivered a peak power of 600 watts to a horizontal dipole whose axis had a north-south orientation. The great circle bearing of the transmitter at the receiver was 88⅓ degrees. Data were taken only during the hours between evening twilight and midnight.

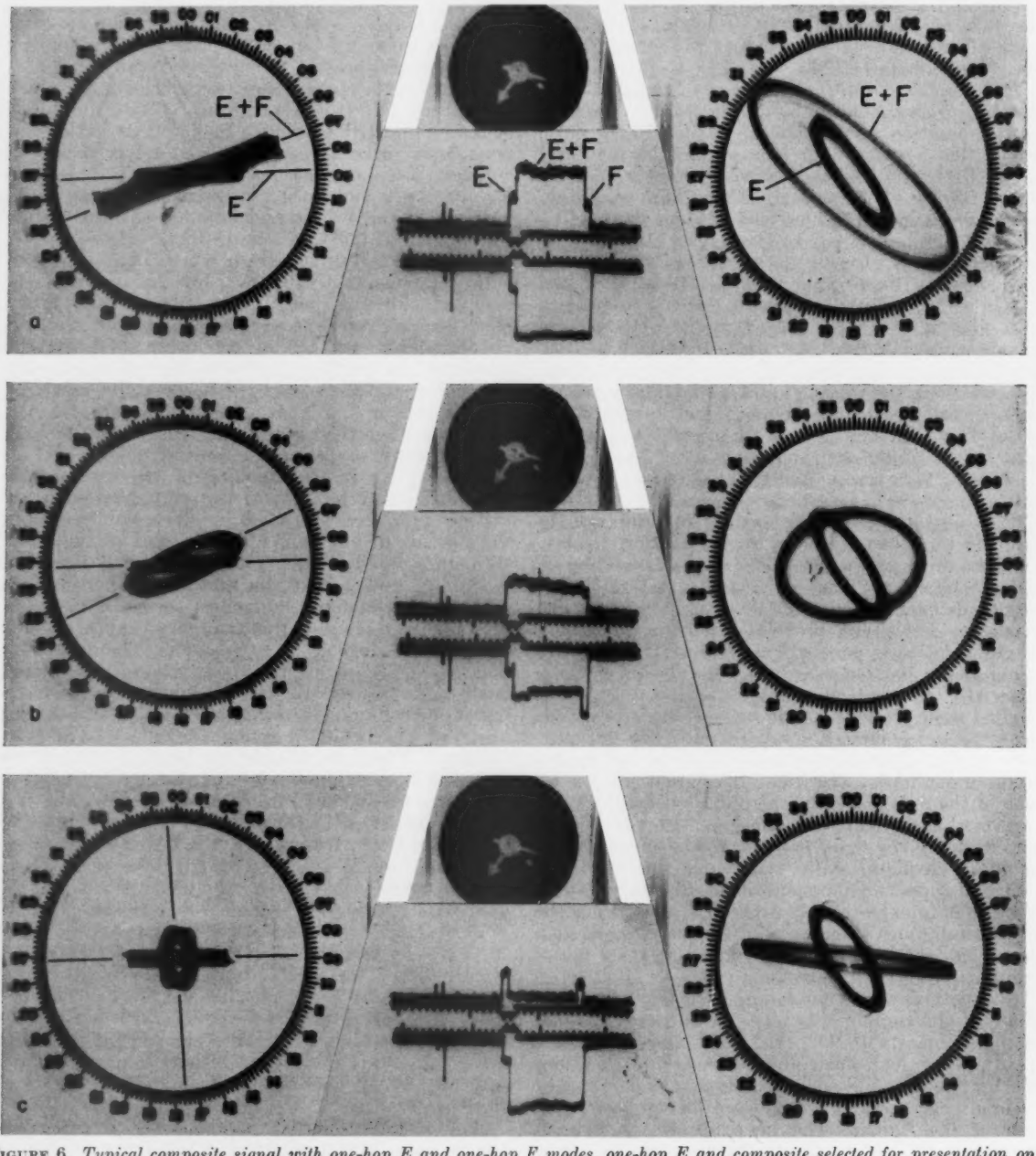


FIGURE 6. *Typical composite signal with one-hop E and one-hop F modes, one-hop E and composite selected for presentation on bearing and polarization displays; successive frames illustrate slip in phase (ϕ) between rays over one fading cycle.*

a, $\phi=0°$; b, $\phi=90°$; c, $\phi=180°$.

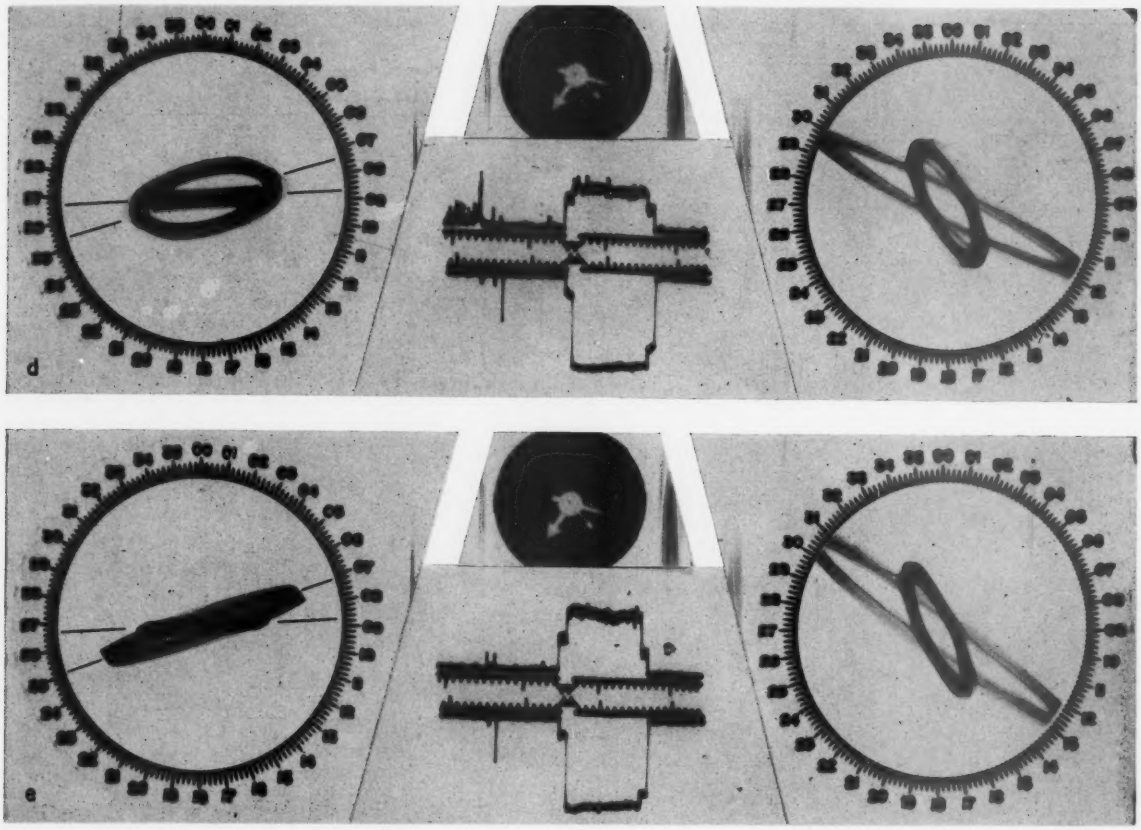


FIGURE 6. *Typical composite signal with one-hop E and one-hop F modes, one-hop E and composite selected for presentation on bearing and polarization displays; successive frames illustrate slip in phase (ϕ) between rays over one fading cycle*—Cont.

d, $\phi=270°$; e, $\phi=360°$.

At the times the data were acquired, the observed MUF for this path for one-hop *F*-layer transmission was above the operating frequency during the entire daily transmission period. The MUF for normal one-hop *E*-layer transmission dropped below the operating frequency shortly after the beginning of the transmission period. Mode identification was accomplished by continuously observing the signal structure pattern throughout the entire late afternoon and evening, from the time when both the normal one-hop *E*- and *F*-layer modes were present, through the period of breakup of the normal *E*-layer mode. The polarization states characteristic of the two magneto-ionic components of the one-hop *F*-layer mode were determined by observing the polarization display generated by the short pulse through that period where the two components interchanged order of arrival as the normal *E*-layer breakup occurred. Only on rare occasions was appreciable splitting of the *E*-layer mode observed.

The significant observed transmission modes are shown in terms of relative time of arrival in figure 10. The times of arrival are all referenced to the one-hop *F*-layer mode (designated C), since that was the only mode regularly present, and always identifiable. No means was available for reference to the time of transmission. In addition to the normal one-hop *F*-layer mode, transmission by a one-hop nighttime *E*-layer mode (designated A) was observed during most evenings, for varying periods of time. In fact all the data recorded on *E*-region signals was obtained on such sporadic transmissions. The mode designated B was tentatively identified as a one-hop reflection from the F_1 region. Its relative time of arrival suits that situation, and its direction of arrival was essentially on-path. The other modes were not positively identified because their occurrence was not sufficiently frequent or prolonged to permit many observations to be made. However, the mode designated *F* was probably a 2-hop *F*-layer reflection.

Bearing information was read from the film recordings, frame by frame, and punched onto paper tape for analysis by the Illiac, the University of Illinois' electronic digital computer. Well over 12,000 in-

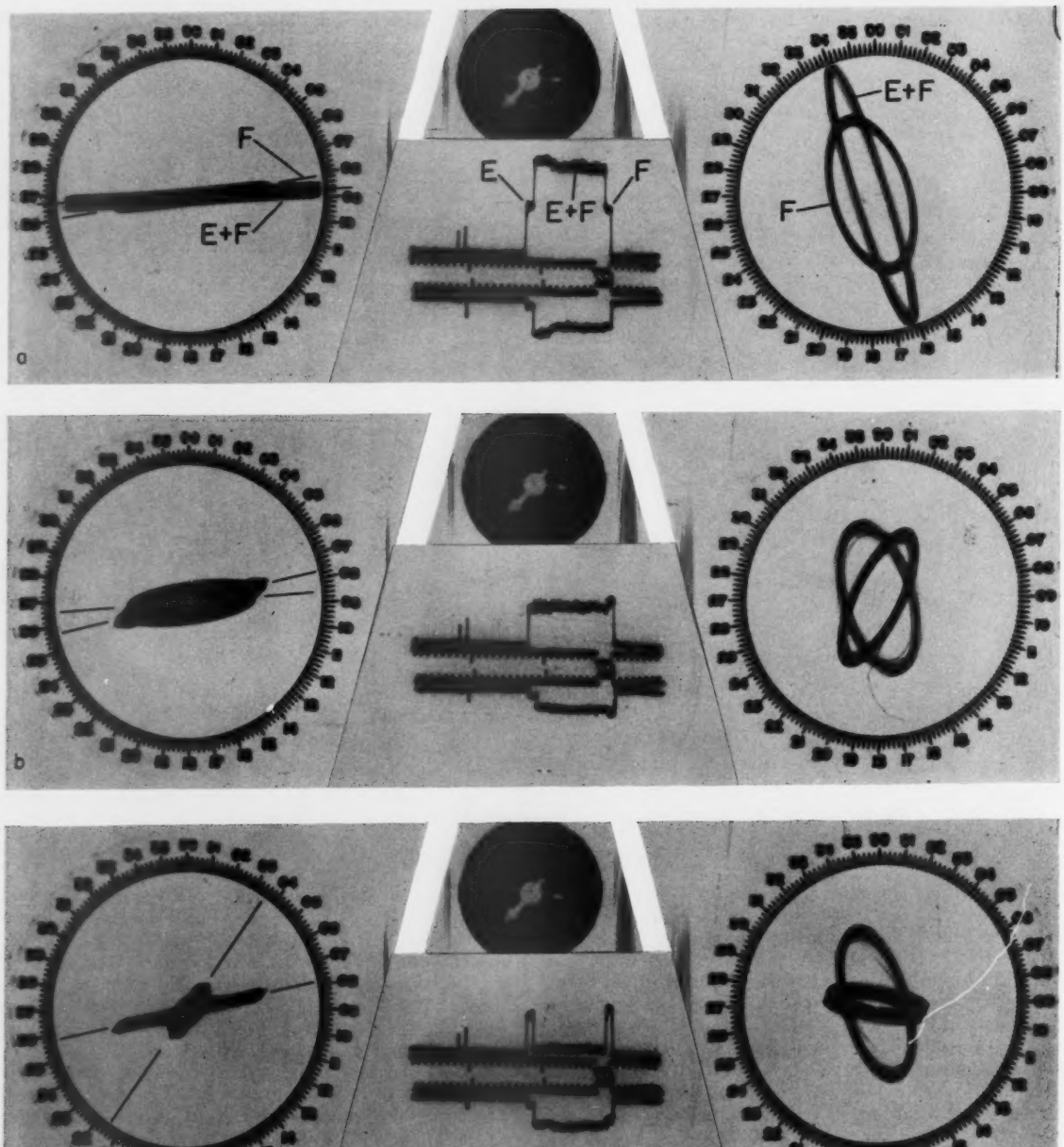


FIGURE 7. *Typical composite signal with one-hop E and one-hop F modes, one-hop F and composite selected for presentation on bearing and polarization displays; successive frames illustrate slip in phase (ϕ) between rays over one fading cycle.*

a, $\phi=0°$; b, $\phi=90°$; c, $\phi=180°$.

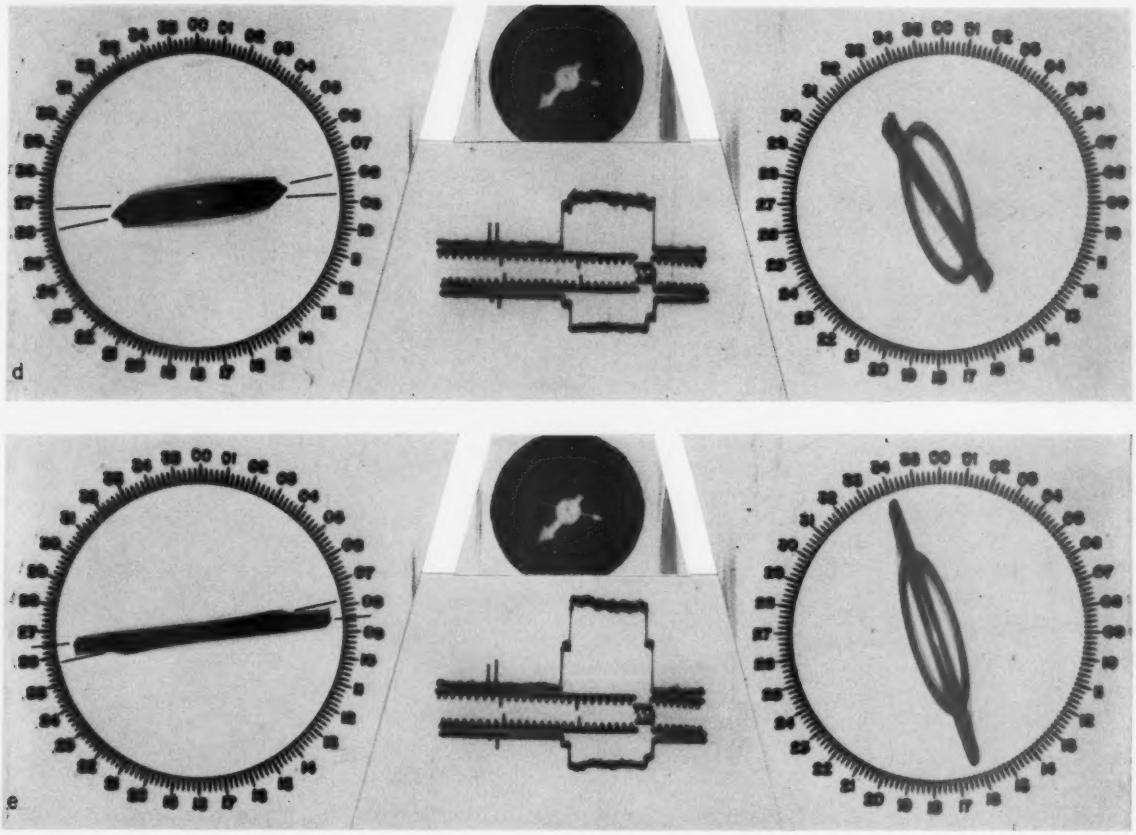


FIGURE 7. *Typical composite signal with one-hop E and one-hop F modes, one-hop F and composite selected for presentation on bearing and polarization displays; successive frames illustrate slip in phase (ϕ) between rays over one fading cycle*—Cont.

d, $\phi=270°$; e, $\phi=360°$.

dividual frames were read. Each sample bearing determination consisted of readings or "snaps" from a series of frames, ranging from 10 to 286, and averaging about 40. Figure 11 gives examples of bearing samples of various types. The range of variability, classed by propagation mode, of the snaps within a sample is shown in figure 12. This is a measure of the rapid (second to second) fluctuation range. Information is given for one-hop propagation via the nighttime *E*-layer, the *F*-layer with ordinary and extraordinary components both present, the *F*-layer with ordinary only, and the *F*-layer with extraordinary only. On the graph bars are marked the minimum, mean, and maximum values of the RMS deviations from the sample mean of the snaps within a sample.

Of interest is the relatively high short-time directional stability of the nighttime *E*-layer mode. This was true in spite of the fact that both magnetoionic rays were present. They were invariably too nearly coincident to resolve. This stability seems indicative of a nighttime *E*-region having structural regularity over a considerable area. The structure must be such that it produces a very nearly specular reflection, rather than a structure which is patchy, thus producing scattered, or nonspecular, reflections. The rather sudden appearance and disappearance of the *E*-layer signal indicate relatively rapid changes in the *E*-layer ion density or height, causing rapid changes in the "skip" distance. Since vertical soundings show little variation in the height of reflection from the *E*-region, ion density variations, having structural uniformity over a fairly large area, seem the more probable explanation.

The large variability in the undifferentiated *F*-layer mode, compared to the single magnetoionic components of that mode, is almost surely due to wave-interference effects (resulting in corrugated phase fronts) between the two magnetoionic rays. The paths traveled by the two magnetoionic rays differ considerably more for *F*-layer transmission than for *E*-layer transmission. This illus-

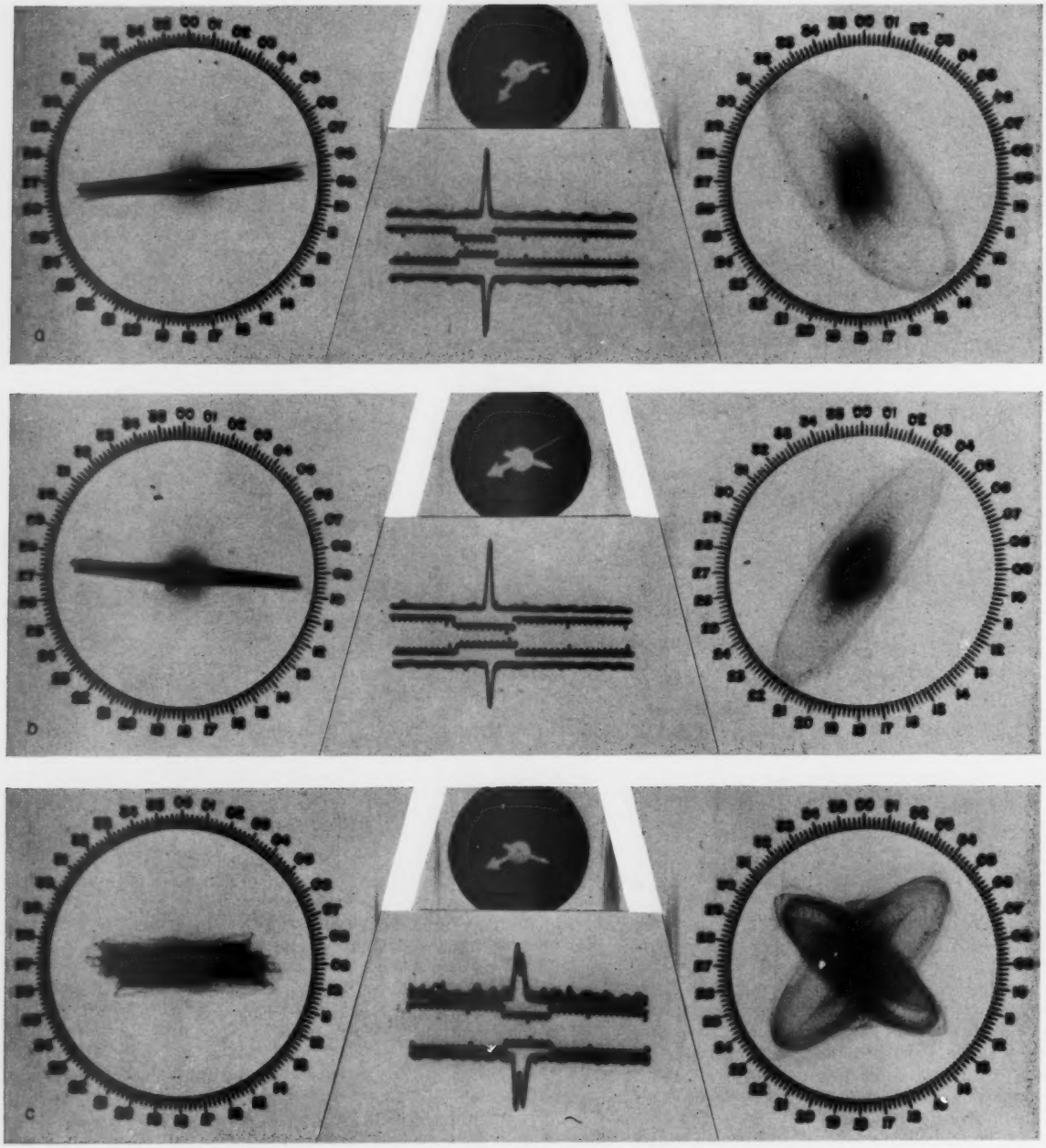


FIGURE 8. *One-hop F mode, short pulses only, expanded sweep.*

a, Ordinary ray only; b, Extraordinary ray only; c, Both rays, essentially resolved.

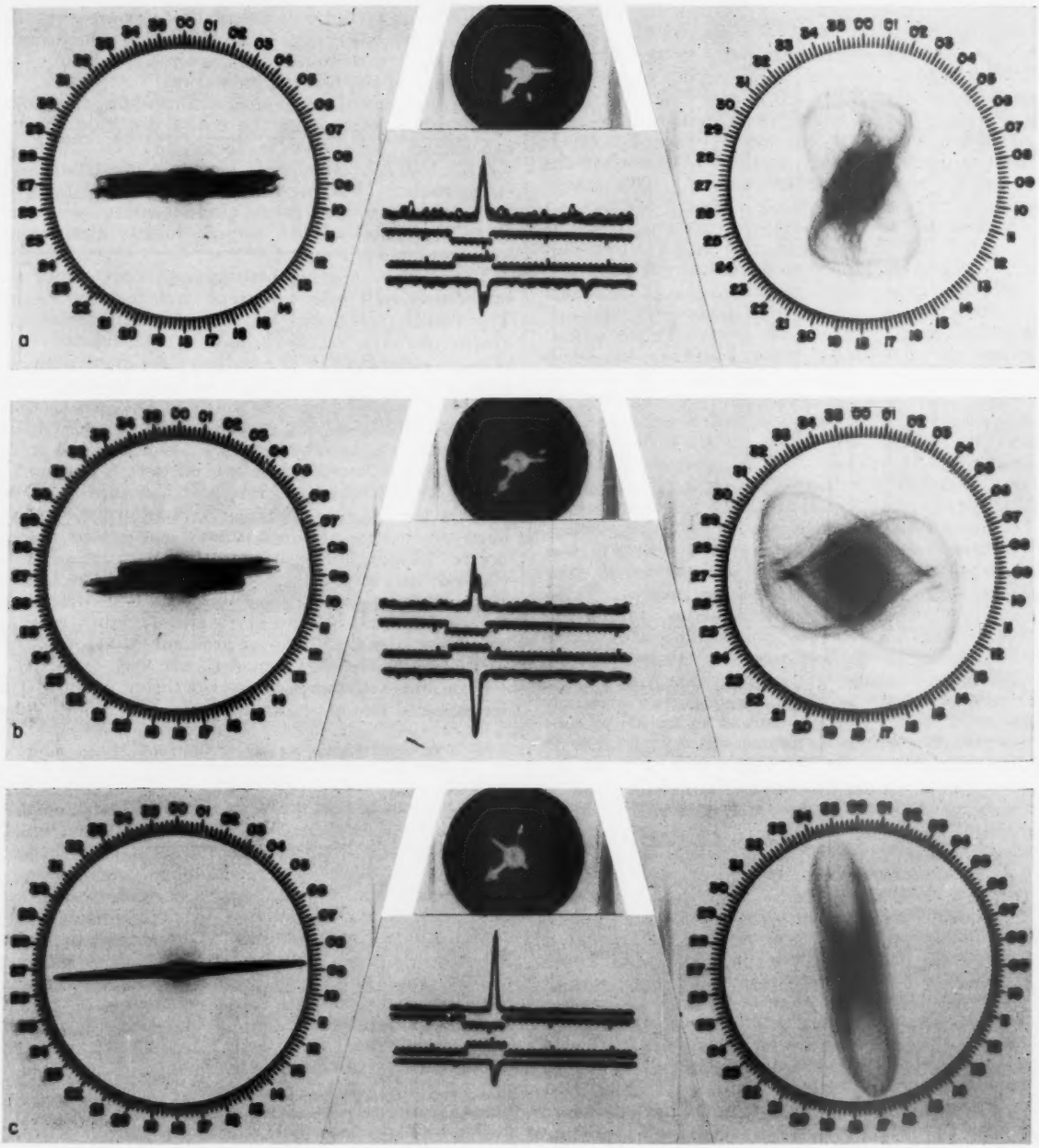


FIGURE 9. *One-hop F mode, short pulses only, expanded sweep, ordinary and extraordinary rays both present, partially resolved.*

a, Vertical components in phase, horizontal out of phase; b, Vertical components out of phase, horizontal in phase; c, Vertical components in phase, horizontal out of phase; two rays nearly coincident.

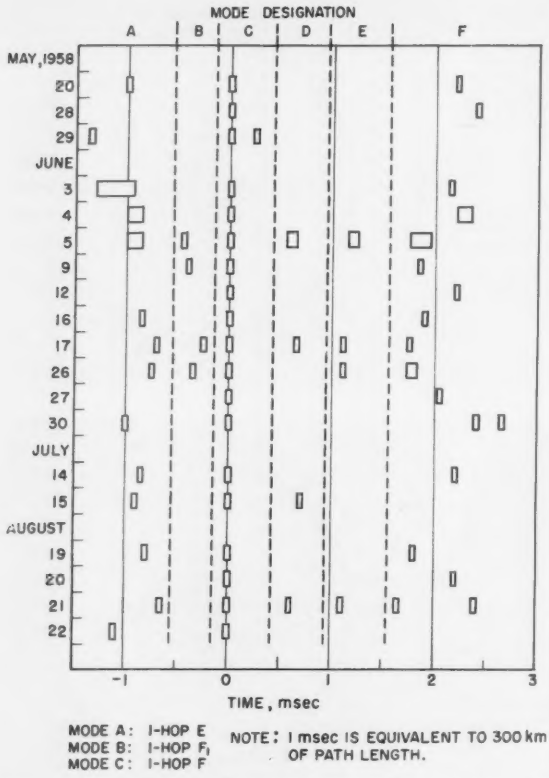


FIGURE 10. *Observed mode structure: note that on a given day not all modes observed were necessarily present simultaneously.*

trates the importance of completely decomposing the signal to its basic rays before drawing conclusions about ray directional characteristics from the variability of the bearing indication.

Distributions of the sample means about the great circle path, taken over the entire duration of the experiment, are shown in figure 13. Notice how tightly the *E*-layer signals cluster about the great circle path. The long-time mean for this mode is only 0.6° from the great circle bearing, and the RMS deviation of the sample means about this long-time mean is only 1.6°.

The *F*-layer bearing distribution chart shows an interesting split near the great circle bearing value. The reason for it can be seen in the distribution graphs for the ordinary and extraordinary components separately. The ordinary component shows a definite bias to the north. In fact, no samples in this mode showed means south of the great circle path. The extraordinary component, on the other hand, shows a definite bias to the south of the great circle path. There are a few entries to the north of this value, but examination of the data in detail reveals that these instances accompany unusually large deviations of the ordinary component to the north. Thus, at these times, both components showed an unusual northerly shift. The Basic Radio Propagation Prediction charts issued monthly by the National Bureau of Standards indicated that at these times a horizontal gradient of ion density, increasing to the north, might have been expected.

Examination of the phase refractive index of the ionospheric region shows that the sense of the bias

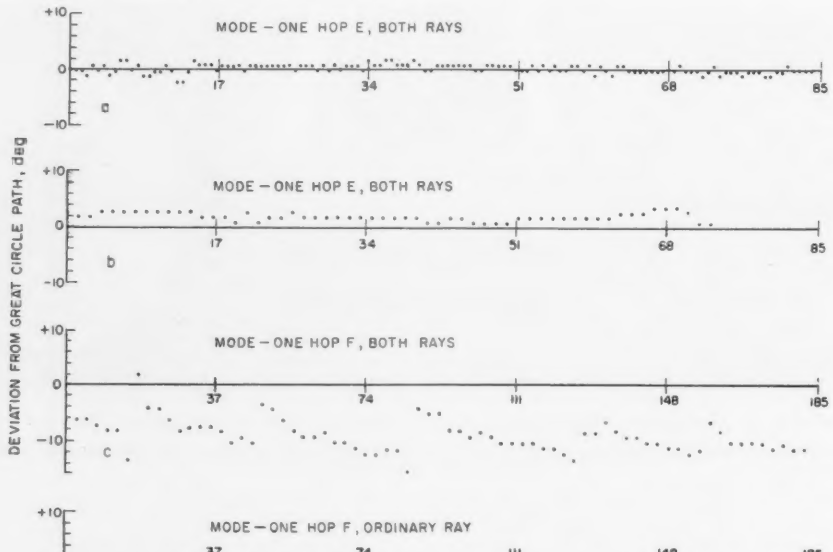


FIGURE 11. *Examples of bearing samples; each sample contains a number of snaps, each snap being read from one frame of film.*

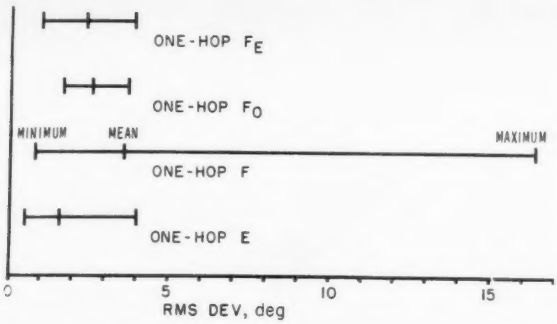


FIGURE 12. *RMS variability of snaps within a sample (short-time variability), classified by mode.*

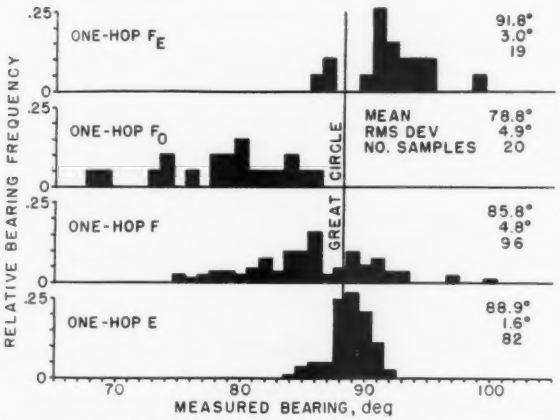


FIGURE 13. *Distribution of sample means about great circle bearing (longtime variability), classified by mode.*

of the two components is as should be expected. No attempt to check theoretically the magnitude of the bias has been made. Notice that the biases of the two components are not equal, the ordinary component bias averaging about 9°, and the extraordinary component bias averaging about 4°. This would be expected as a consequence of the tilt of the earth's field at this latitude. Note also that the average of the means of the ordinary and extraordinary components taken separately is only 0.5° from the mean for the undifferentiated mode (the data for the two magneto-ionic components separately is not included in that for the composite mode). Thus if directional measurements on signals in this mode were taken over a period of time, they would show a northward bias. If a position fixing task permitted long-time measurements, such a bias might be taken into account, thereby improving the fix accuracy.

3. Ray Resolution by Direction-of-Arrival

When the component rays of a multimode signal arrive from directions which differ sufficiently, they can be resolved by a directive antenna system and displayed separately. The Wullenweber direction-finding system at the University of Illinois is such a system. It consists of an array of 120 vertical antenna elements (folded monopoles) disposed on a circle 994 feet in diameter on the outside of a reflecting wire screen 955 feet in diameter. The elements are used 40 at a time, those in use being coupled to the receiver through a phasing system which forms a fan-shaped radiation pattern with a small azimuthal beamwidth. A rotary capacitive switch is provided to scan the directive pattern in azimuth by successively connecting a new element at one end of the set of 40, while dropping one at the other end. The capacitive switch is motor driven at a rate of 900 rpm. Thus a complete azimuth scan is obtained in $\frac{1}{15}$ sec. The switch can also be manually set.

The output signal from the receiver is visually displayed, causing a radial deflection of the spot on a cathode ray tube whose deflection yoke rotates in sychronism with the scanning switch.

The system is designed to operate in the frequency range from 4 to 16 Mc/s. The half-power beamwidth ranges from 14° at 4 Mc/s to 3° at 16 Mc/s.

The signals from radio station WWV are particularly convenient for study because of their simplicity and their continuous presence on several frequencies, and because of the favorable transmitter location with respect to the receiving site. The direct path length is about 1000 km. The variety of propagation phenomena whose directional characteristics are observable is gratifying, and not a little surprising. The system has been particularly successful in conveying a description of the angular spectrum of signals which are too complex to be handled on smaller direction finding systems. Some examples of such signals are described in the following paragraphs. In interpreting the figures, note that the great circle bearing of the transmitter is 93°.

One of the interesting phenomena observed is the nature of the signal breakup as the MUF drops down through the operating frequency. This is presently observable on the 10 Mc/s transmissions of WWV almost every evening. Figure 14 shows a sequence of photographs of the azimuth display taken during an afternoon and evening. The first in this series shows the nature of the display on a single-ray signal. It is essentially a polar plot of the array radiation pattern for this frequency (10 Mc/s). The irregularity in the nose of the pattern is a commutation defect. Multiple ray arrivals at angles appreciably off the great circle bearing are evident in the intermediate photographs. Multiray character can be observed even where the rays are not resolved by the broadening of the response pattern. In the last photograph only scatter signal is present. It is, however, appreciably above the noise level. While the details of the signal behavior vary from evening to evening, the general pattern is usually similar to that shown here.

In figure 15 are shown two examples of fairly strong, but highly non-specular incident signals. They are from WWV on 10 Mc/s. Such signals are characterized by very high fading rates (up to 20 or 30 fades per second).

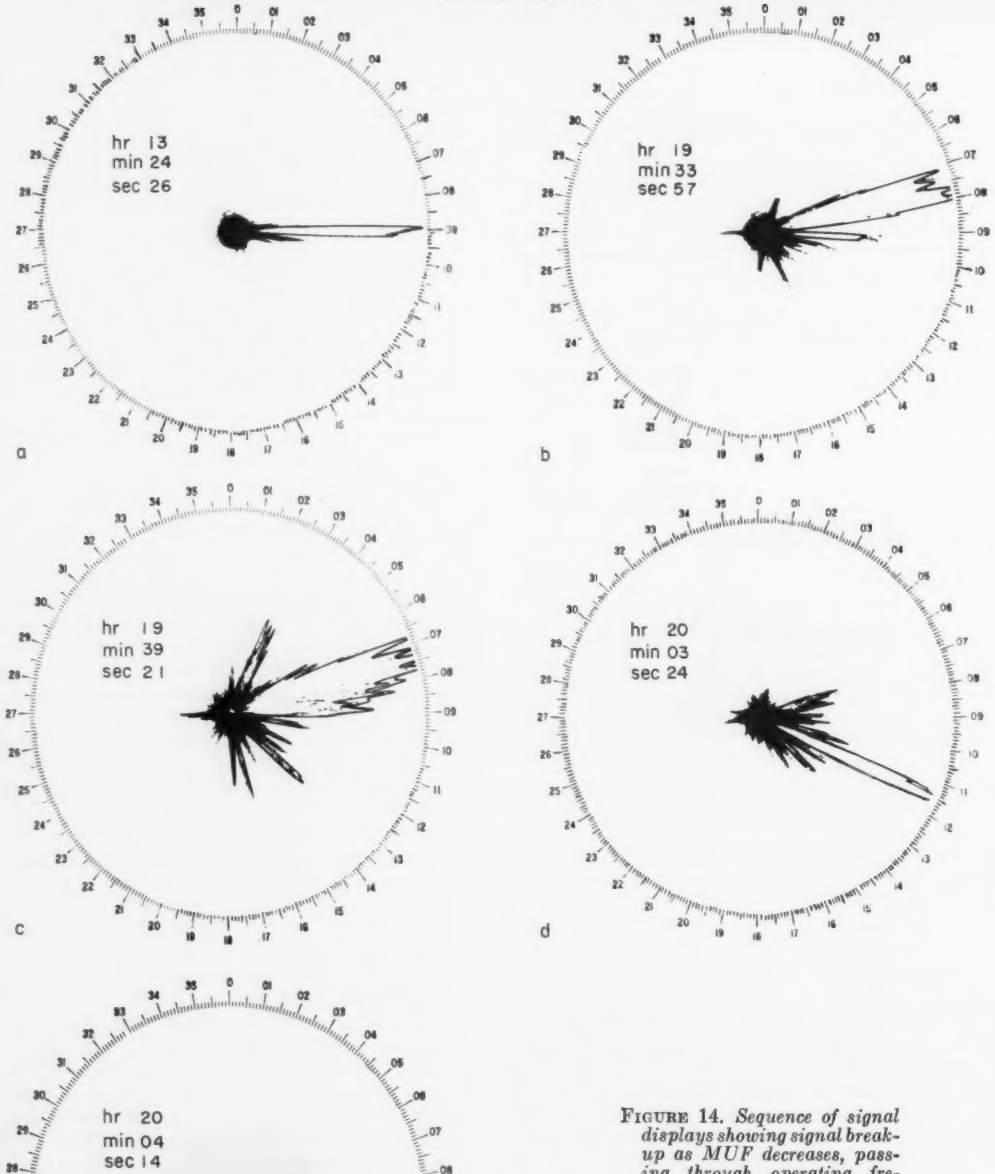


FIGURE 14. *Sequence of signal displays showing signal breakup as MUF decreases, passing through operating frequency; signal is from WWV on 10 Mc/s, great circle bearing is 93°.*

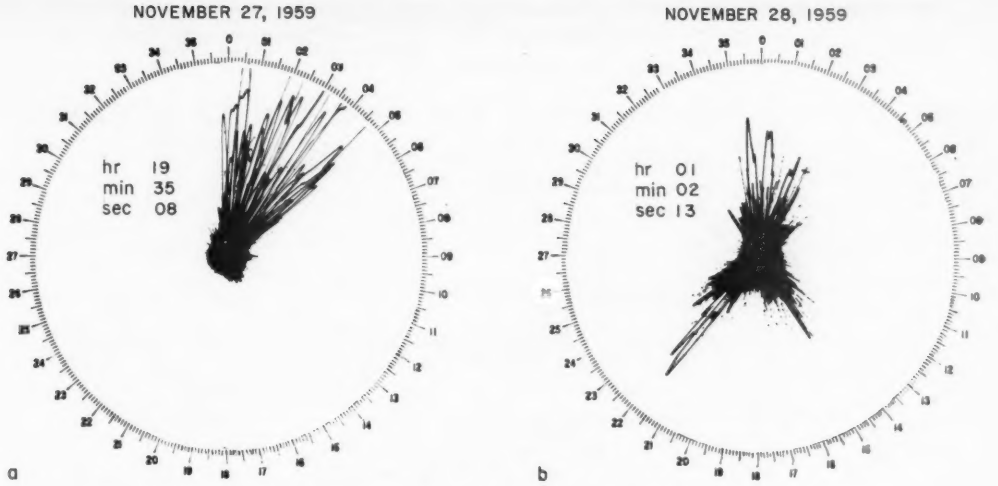


FIGURE 15. *Examples of anomalous signals propagated via spread-F and scatter structures.*

a, Spread-*F* signal; b, Scatter signal.

Signals of the type shown in figure 15a seem to be associated with the existence of spread-*F* over eastern Canada. Such displays are frequently observable, and may last from a few minutes to several hours. Energy may be incident over an angular sector ranging from 5° to 60°.

Using some reasonable assumptions, it has been possible to estimate the point of reflection for some signals of this type. It is often possible to receive a normal on-path signal from WWV on 5 Mc/s at times when anomalous displays are observed on 10 Mc/s. By setting the steerable beam manually, it has been possible to measure, as a function of direction of ray arrival, the difference in time of flight between the time ticks propagated in a regular mode and that of the ticks propagated in the anomalous mode. Assuming that the anomalous reflection occurs at *F*-region heights, knowledge of the direction of arrival and excess time of flight permits location of the reflection point. A number of such determinations are shown in figure 16. Each series of points connected by a line represents a number of measurements on one display of the phenomenon. A period of the order of 20 minutes was required to obtain the measurements in each set. The different sets were measured on different occurrences of the phenomenon. A number of parallels of geomagnetic latitude have been added to the map for reference to show the tendency of the reflection point loci to lie, at any one time, along parallels of geomagnetic latitude.

Figure 15b shows another type of non-specular signal having even less regularity and higher fading rate than that of figure 15a. Such signals have been called "scatter" signals. There are times when the individual raylet behavior in these signals is suggestive of reflection from dense meteor trails. Attempts have been made to plot reflection point loci for signals of this sort. However, because of the very rapidly changing configuration of the display, these attempts have met with much less success than was achieved on signals such as shown in figure 15a.

All of these phenomena are being made subjects for study in greater detail.

4. Conclusion

The two experimental studies described demonstrate the power of direction-finding techniques in studying propagation phenomena. A particularly promising extension of this work is the tracking of a frequency-swept ionosonde with high-resolution DF equipment. An experiment of this type is being planned.

The experiments described above were performed by members of the staff of the Radio Direction-Finding Research Laboratory, Department of Electrical Engineering, University of Illinois, as parts of research programs sponsored by the Office of Naval Research and the Bureau of Ships.

5. Bibliography

Bailey, A. D., An investigation of the direction of arrival of radio waves, Thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, University of Illinois, Urbana, Ill. (1954).

Bain, W. C., The calculation of wave-interference errors on a direction-finder employing cyclical differential measurement of phase, Proc. Inst. Elec. Eng. **100**, Pt. III, pp. 253–261 (September 1953).

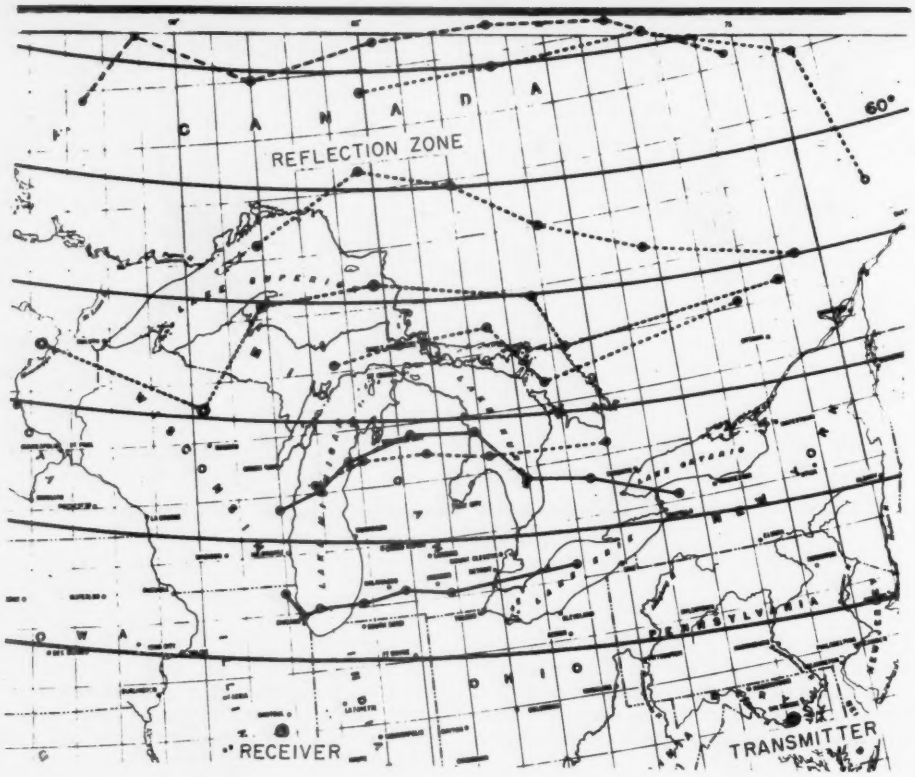


FIGURE 16. *Estimated reflection points for spread-F signals.*

Bain, W. C., On the rapidity of fluctuations in continuous-wave radio bearings at high frequencies, Proc. Inst. Elec. Eng., **102,** Pt. B, pp. 541–543 (July 1955).

Bain, W. C., The theoretical design of direction-finding systems for high frequencies, Proc. Inst. Elec. Eng. **103,** pt. B, pp. 113–119 (January 1956).

Bain, W. C., Possible errors of a particular wide-aperture direction-finder, Proc. Inst. Elec. Eng., Mono. No. 170R, pp. 1–12 (March 1956).

Bain, W. C., Fluctuations in continuous-wave radio bearings at high frequencies. Proc. Inst. Elec. Eng. **103,** p. 560 (July 1956).

Barfield, R. H. and W. Ross, The measurement of lateral deviation of radio waves by means of a spaced loop direction finder, J. Inst. Elec. Eng. **83,** p. 98 (1938).

Bramley, E. N., Some aspects of the rapid directional fluctuations of short radio waves reflected at the ionosphere, Proc. Inst. Elec. Eng. **102,** pt. B, pp. 533–540 (July 1955).

Barmley, E. N., Some comparative directional measurements on short radio waves over different transmission paths, Proc. Inst. Elec. Eng. **102,** pt. B., pp. 544–549 (July 1955).

Bramley, E. N., H. F. bearing variations of an Adcock Direction-Finder, Proc. Inst. Elec. Eng., Mono. No. 175R, pp. 1–7 (April 1956).

Clemmow, P. C. and R. F. Mullaly, The dependence of the refractive index in magneto-ionic theory on the direction of the wave normal, The Physics of the Ionosphere, Report of the Physical Society Conference on the physis of the ionosphere held at the Cavendish Laboratory, September, 1954, The Physical Society (London, 1955).

Gleason, R. F. and J. H. Trexler, Ionospheric limitations in the ultimate accuracy of direction finding, NRL Memorandum Report No. 61, Naval Research Laboratory (Washington, D.C., 1952).

Heiligtag, Th., Uber die Grunde der Missweisungen beim Richtungempfang, Jahrbuch Zeitschrift fur Drahtlose Telegraphie and Telephonie **21,** p. 77 (1923).

Jordan, E. C., et al., Summary technical report, Technical Report No. 4, The Radio Direction Finding Research Laboratory, Department of Electrical Engineering, University of Illinois, Urbana, Ill.

Ross, W., Lateral deviations of radio waves reflected at the ionosphere, Department of Scientific and Industrial Research, Radio Research, Special Rept. No. 19, H. M. Stationery Office (London, 1949).

Ross, W., Fundamental problems in radio direction finding at high frequencies, J. Inst. Elec. Eng., pt. IIIA **94,** pp. 154–165 (1947).

Smith, S. B., and H. G. Hopkins, H. F. direction finding, Wireless Engineer, **31,** pp. 11–14 (January 1954).

Fundamental Principles of Ionospheric Transmission, Department of Scientific and Industrial Research, Radio Research, Special Rept. No. 17, H. M. Stationery Office (London, 1948).

Ionospheric Radio Propagation, NBS Circ. 462 (1948).

(Paper 65D3–120)

JOURNAL OF RESEARCH of the National Bureau of Standards—D. Radio Propagation
Vol. 65D, No. 3, May–June 1961

Diversity Effects in Long Distance High Frequency Radio Pulse Propagation[1, 2]

S. A. Bowhill [3]

(August 15, 1960; revised January 4, 1961)

Spaced antenna measurements are described, made on a 8,600-km path between Colombo, Ceylon, and Great Baddow, England, using pulsed radio signals. Simple interpretations in terms of E and F region multiple reflections give good agreement with observed delays for the various echoes.

The correlation between echo amplitudes at spaced antennas was found to be much greater for a pulsed signal than for a CW signal, indicating that most of the diversity in long-distance CW transmission arises through phase incoherence between the various orders of reflection.

1. Introduction

A study of the ionospheric irregularities responsible for the fading of long distance radio signals can best be accomplished by oblique incidence pulse measurements. Previous workers [Wilkins and Kift, 1957: Silberstein, 1958] have shown that very complex modes of propagation may be present on long-distance paths. Detailed theoretical study of these modes has been lacking until recently, when Kift (1960) gave a very detailed method for studying the entire propagation path.

This paper describes the basic phenomena involved, and gives some experimental results obtained on 16.16 Mc/s frequency, using transmissions from Colombo, Ceylon, received at Great Baddow, England.

The complicated pattern of the pulse signal received led to a consideration of possible reflection mechanisms; third and other order F region reflections were found, with E region multiples present on occasion. The complexity of a typical pattern implies that it is a difficult matter to distinguish, in the CW case, between fading due to phase interference between different orders of reflection, and fading of the reflections themselves.

The path differences for the various modes are calculated in section 2, and the experiments are described in section 3. Section 4 compares the calculated path differences with those experimentally observed; the fading analysis is given in section 5.

[1] The work described in this paper was performed by the author while at Marconi's Wireless Telegraph Co., Ltd., Baddow Research Laboratories, Great Baddow, Essex, England.

[2] Paper presented at the Conference on Transmission Problems Related to High-Frequency Direction Finding, at UCLA, June 21–24, 1960.

[3] Ionosphere Research Laboratory, Pennsylvania State University, University Park, Pa.

2. Ray Theory of Long-Distance Propagation

A radio wave traveling a long distance between points on the earth's surface, such as A and M (fig. 1), may be reflected sometimes by the E layer and sometimes by the F layer. In section 2.1, the elementary theory (ignoring ionospheric tilts) for a thin layer is given; in section 2.2, this is extended to include a sporadically occurring layer at a lower height than the main reflection. Section 2.3 describes how a layer of finite thickness may be quickly allowed for.

2.1. Single Thin Layer

For any particular path, the distance d of the receiver from the transmitter is fixed, and the only variables needed to determine all the properties of a particular mode of propagation are the height of the reflecting layer, h, and the order n of the reflection, i.e., the number of "hops" the radio wave makes from the earth to the layer and back over its path. If the angle of arrival, the angle the wave makes

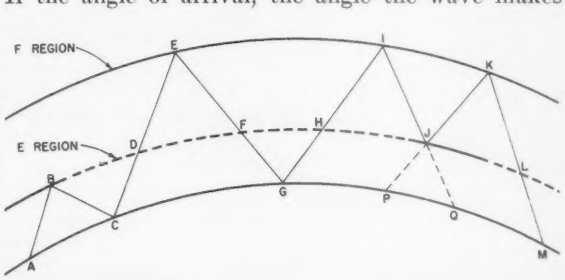


FIGURE 1. *Typical reflection geometry.*

with the horizontal at the earth's surface, is θ, one may write the relations

$$d=R\sin(d/2nR)\tan(\theta+d/2nR)-R+R\cos(d/2nR), \quad (1)$$

$$p=2nR\sin(d/2nR)\sec(\theta+d/2nR)-d, \quad (2)$$

where p=path difference between the reflected wave and a wave traveling directly over the earth's surface.

The eqs (1) and (2) may be regarded as a parametric equation for p in terms of h.

Numerical application of the formulas gives the set of curves shown in figure 2. Three sets of curves are shown on this graph. The first set, for integral values of n, the order of reflection, shows the variation of the time delay p with the height of the reflecting layer, and it is evident that the greater the height of reflection, the greater will be the spacing between successive orders of reflection in the pulse pattern.

Not all the orders of reflection shown are possible in any practical case. Firstly, if we neglect diffraction effects, it is not possible for radio waves to arrive at the receiver with effective angles of arrival less than zero. A set of curves has therefore been drawn showing the angles of arrival of the different orders of reflection at various layer heights. These curves of constant angle of arrival happen to be nearly straight lines, and also are approximately parallel. This implies a relation between the relative time delay of an order of reflection and its angle of arrival, practically independent of the height of reflection. The upper limit is determined by the electron concentration of the ionized layer. The greater this concentration, the steeper the angle at which a radio wave may be incident upon it and still be reflected. This is expressed by the relationship,

$$f=f_0 \sec i,$$

where f=radio wave frequency,
i=angle of incidence of wave at layer=$(n/2)-\theta-d/2nR$,
f_0=vertical incidence critical frequency of the layer.

As the graph is constructed for a particular frequency, namely 16.16 Mc/s, curves of constant f can be drawn in on figure 2. If the vertical incidence critical frequency of the layer is known, the highest permissible order or reflection may be directly read off.

It is now necessary to compare these theoretical suggestions with actual ionospheric data observed at vertical incidence pulse sounders for the period in question. The path from Colombo to Great Baddow passes, in the main, over a region where there are no pulse sounders whose results are regularly available. However, use may be made of the control-point method, which is normally applied to predicting the behavior of long-distance HF propagation paths.

The western control point is fairly near the ionosphere station at Graz, Austria, while the eastern point is at the same latitude as Bombay, and not too far removed in longitude. Mean ionospheric data for the month of July, 1953 are available from these two ionospheric sounding stations, and are given in table 1, with allowance for the difference between G.m.t. and the local solar time at the control

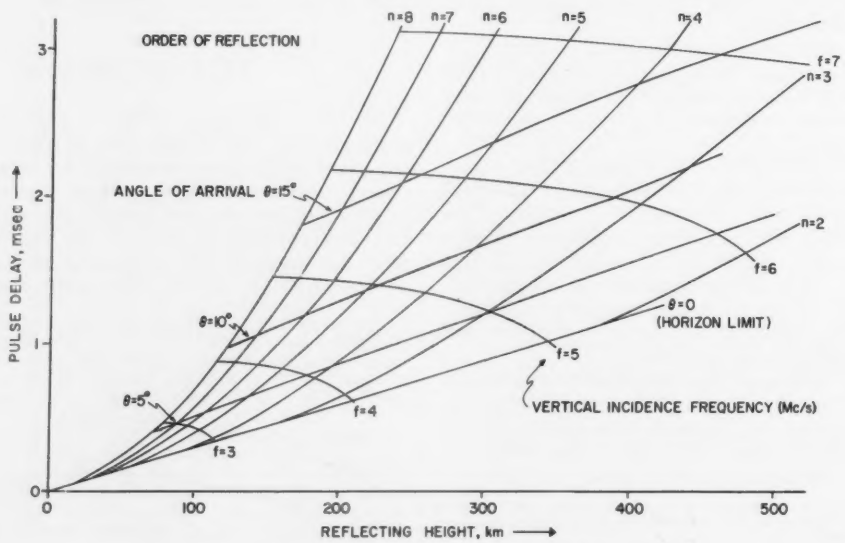


FIGURE 2. *Colombo-Baddow path delays.*

TABLE 1

Time G.m.t.	E. control point		W. control point		Mean virtual height
	$f^\circ F2$	$h' F2$	$f^\circ F2$	$h' F2$	
	Mc/s	*km*	*Mc/s*	*km*	*km*
08	9.0	433	4.9	300	366
09	9.3	432	5.0	315	374
10	9.6	419	5.2	300	360
11	9.5	408	5.2	340	374
12	9.3	383	5.1	310	346
13	8.8	361	5.0	330	346
14	8.1	327	5.0	320	324
15	7.0	324	4.9	300	312
16	5.9	316	5.0	310	313
17	4.8	311	5.0	280	296

point. It is clear that the ionization density at the western control point is the limiting factor, except at 1700 G.m.t.

Corresponding values of the effective reflection height and limiting frequency for the western control point have been plotted in figure 3 for each hour in the afternoon. If a line is drawn vertically downwards from the point corresponding to any hour, it should intersect the curves of constant n at the experimentally observed values of delay. This will be compared with experimental data in section 4.

2.2. Multiple Thin Layers

In this section a simple graphical method will be described for deriving the theoretically expected pulse pattern if multiple reflections are present. It will be assumed that the signal is propagated by reflection at two sharply bounded layers, situated at known heights. The upper layer *EIK* on figure 1, will be assumed uniform in properties, while the lower layer *DFHJL* will be taken to be inhomogeneous in the horizontal direction. At *B* and *J* it is present, and reflects the ray; at *D*, *F*, *H*, and *L* it is absent, and the ray passes through the layer without deviation. The reflection at *B* is an "*E*" reflection; that at *E* is an "*F*" reflection; while that at *I*, *J*, and *K* will be termed an "*M*" reflection.

Two points must be noted about this type of propagation. The first is that, provided the numbers of each type of reflection remain the same, they may be permuted in any order without affecting the characteristics of the received signal. The second is that the angle of departure of signal at *A* is equal to the angles of arrival at *C*, *G*, and *M*, and also determines directly the angles of incidence of the ray at the two layers, since they are at predetermined heights.

The characteristics of the *M* mode of propagation are closely related to the *E* and *F* reflections. In fact, as can be seen from figure 1, if the extension lines *JP* and *JQ* are drawn, its transmission distance *GPQM* is equal to that of two *F* reflections, less that of one *E* reflection. The total transmission distance *GHIJKLM* is also equal to that of a $2F$ reflection less that of a $1E$ reflection. So *M* reflections will not be treated separately, but simply included as $2F+\bar{1}E$ reflections, where $\bar{1}$ denotes a negative order of reflection. For instance, the path illustrated in figure 1 is just equivalent to $1E+1F+2F-1E=3F$.

If propagated solely by reflection at the *E* layer or the *F* layer, let a ray make N_E or N_F hops, respectively, in traveling a distance d (the fact that N_E and N_F may be fractional presents no analytical difficulty; they may be regarded as obtained by interpolation between integral values of N_E and N_F for the particular value of θ). The ray therefore travels a distance d/N_E in one *E* hop, and d/N_F in one *F* hop. If the ray makes mE reflections and nF

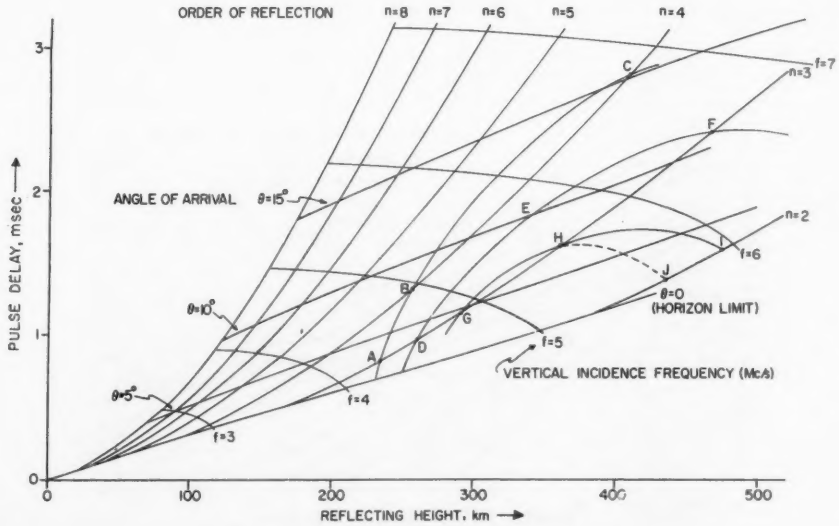


FIGURE 3. *Thick layer delays for July 3, 1953.*

reflections, clearly

$$\frac{m}{N_E}+\frac{n}{N_F}=1. \tag{3}$$

It is convenient to represent the various orders of reflection on a graph like figure 4. Each intersection of integral ordinates and abscissas represents a possible mode of propagation. Points along the horizontal axis represent rays propagating by 1, 2, 3, etc., reflections at the E layer alone; similarly, on the vertical axis they represent reflections occurring at the F layer alone. Points not lying on either axis represent mixed modes of propagation, partly E and partly F.

Now eq (3) is just the equation of a straight line on this diagram, since N_E and N_F are constant for constant θ. The full lines on figure 4 are for $\theta=0°$, 5°, 10°, 15°, and 20°, for a height of reflection of 120 km for the E layer, and 250 km for the F layer.

The points on the negative side of the E axis represent the M type reflections mentioned previously; for instance, the point P on figure 4 is $6F+\overline{3}E$, or 3 M reflections. A line has been drawn from the origin with a slope of -1; no intersections on or below this have any physical meaning.

One may easily find the total time delay of a mixed order of reflection. If the ray would be delayed by p_E or p_F when propagated entirely by reflection at the E or F regions respectively, then its delay for unit distance is p_E/d and p_F/d; the delay experienced by an $mE\ nF$ echo is then

$$p=p_E\cdot\frac{m}{N_E}+p_F\cdot\frac{n}{N_F}. \tag{4}$$

The loci of constant p on figure 4 are therefore again straight lines. They are drawn as dashed lines, for delays of 0.5, 1, 1.5, 2, 2.5, and 3 msec.

The highest order of F region reflection which can be propagated is limited by the F region critical frequency; the relevant frequency for each angle for an F region height of 250 km is inserted opposite the F axis on figure 4.

Since the nature of Es ionization is to occur in relatively intense patches, it will be assumed that the E mode is determined by its spatial extent, rather than its critical frequency as such. From eq (3), E propagation occurs for m/N_E of the path; if it is assumed that there must be on an average this amount of Es to support the $mE\ nF$ mode of propagation, lines of constant fractional coverage of Es can be inserted on figure 4; they are shown for fractions of 0.2, 0.4, 0.6, and 0.8.

As an example, one may take an F critical frequency of 6 Mc/s and a fractional coverage of 0.2. All the points in the quadrilateral $ABCD$ will be possible orders of reflection; they are

$$3F,\ 4F,\ 5F$$
$$1E\ 3F,\ 1E\ 4F,\ 1E\ 5F$$
$$\overline{1}E\ 4F,\ \overline{1}E\ 5F,\ \overline{1}E\ 6F$$
$$\overline{2}E\ 6F.$$

The delays of these various orders can be interpolated between the straight lines of constant delay.

It is evident that when Es propagation becomes important each order of F reflection will have a series of E multiples associated with it, one set arriving earlier, the other set later, at the receiver. The interval between these successive orders of E reflection varies from 0.15 msec for the $3F$ reflection to 0.3 msec for the $7F$ reflection. If the Es ionization is present over a large proportion of the path, the E multiples associated with successive orders of F reflection overlap, giving a complex pattern.

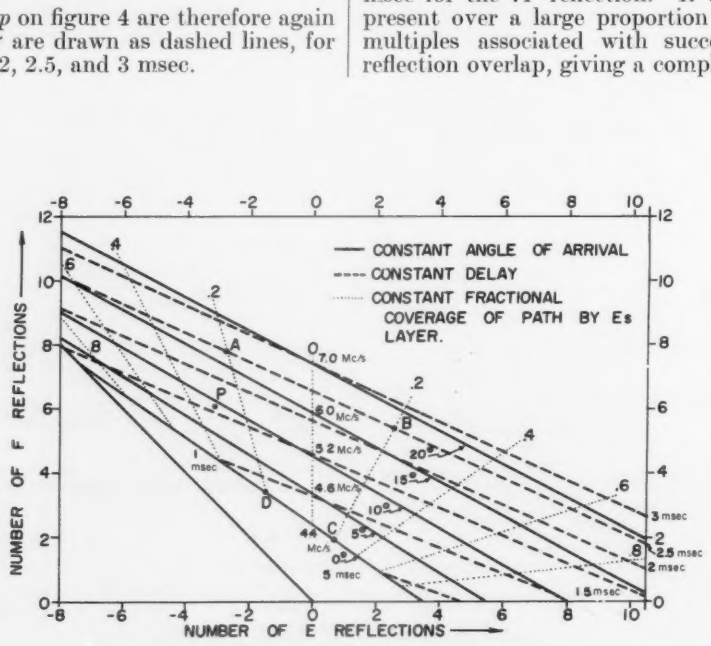


FIGURE 4. *Statistical effect of* Es *coverage.*

2.3. Extension to Thick Layer

Assuming the Breit and Tuve theorem to hold, the finite thickness of the reflecting layer may be allowed for by plotting the $h'f$ curve on figure 3, using the limiting frequency scale derived previously. The curve ABC on figure 3 represents the $h'f$ curve of a parabolic region, with the height of maximum ionization, h_m=300 km, semithickness 120 km, and critical frequency 7 Mc/s. The delays and angles of arrivals of the various orders of reflection which would be observed, if this layer were present throughout the path, are given by the intersections of this curve with the n=3, 4 etc., lines. In the limit of a very thin region, the curve ABC becomes a vertical line, terminating sharply at the critical frequency.

It should be noted that the curve ABC, if extended beyond the right edge of figure 3, will intersect the n=4 line again, at a higher angle of arrival. This, of course, represents the so-called "high ray" sometimes observed on oblique incidence paths.

3. Experimental Measurements

3.1. General Description

The experiments to be described used transmissions from an Air Ministry sender at Colombo, Ceylon, on a frequency of 16.16 Mc/s. The duration of each transmitted pulse was 100 μsec and a rhombic transmitting aerial was used, the polar diagram of which (see fig. 5) had a maximum at about 10° to the horizontal. The experiments commenced on the pulse transmissions at Baddow in April 1953, and continued until the end of January 1954. Apart from unavoidable breaks in transmission, the signals were sent out during daylight hours on one day every week. At the end of January 1954, some special transmissions were arranged by the Radio Research Board on four consecutive days. The cooperation of the Radio Research Board and the Air Ministry throughout the experiments proved most valuable.

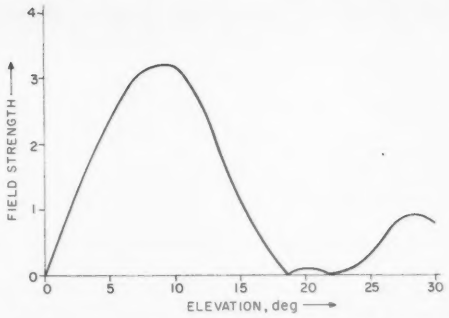


FIGURE 5. *Transmitting antenna polar diagram.*

3.2. Receiving Arrangements

The signal pulses were transmitted at a rate of 25/sec; by the time they reached the receiver, the presence of different propagation modes had lengthened them to about 1 msec. Consequently, some form of triggered time-base was necessary to display the incoming waveforms on a cathode-ray tube screen. Various methods using the signal itself to trigger the time-base proved unsuccessful since the signal frequently faded below local and atmospheric noise. As the source of 25 c/s at the transmitter was a highly stable crystal oscillator, a similar oscillator was later used at the receiving station.

The block diagram of the receiving equipment is shown in figure 6. The signals were received on a half-wave dipole antenna by a CR150 communication receiver, working on its widest bandwidth of 13 kc/s. Under some circumstances it was necessary to use a bandwidth of only 8 kc/s, due to the presence of powerful CW stations on nearby frequencies; however the difference in the observed pulse waveforms when these two bandwidths were used alternately, suggested that little information was lost by using the 8 kc/s bandwidth. The receiver output was displayed on an oscilloscope; the 25 c/s triggering impulses for the oscilloscope time base were derived

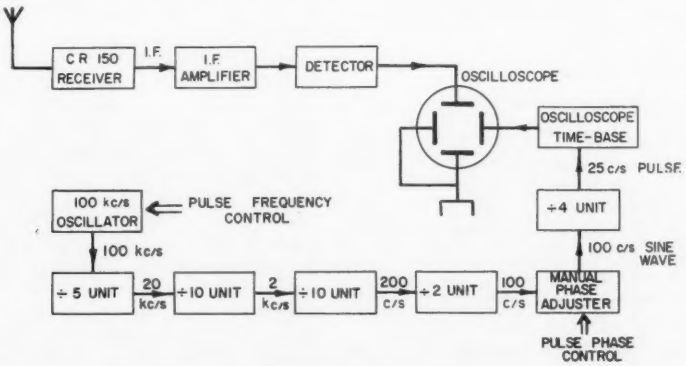


FIGURE 6. *Equipment block schematic.*

from a 100 kc/s crystal oscillator, the frequency of which could be altered by about 1 part in 4,000 by means of a frequency pulling circuit.

Two typical waveforms (instantaneous still photographs of the cathode ray tube display) are shown in figures 7(a) and 7(b). These photographs are two of the series described in section 5.2. It is seen that the length of the pulse train is over 1 msec, and that it may be quite complex.

3.3. Types of Echo Observed

To study the structure of the echo in detail, a method of recording was used in which the incoming signal voltage modulated the brightness of the oscilloscope trace. A photographic film, moved at right angles to the trace, then gave a record similar to the group-height-time records familiar in vertical incidence work.

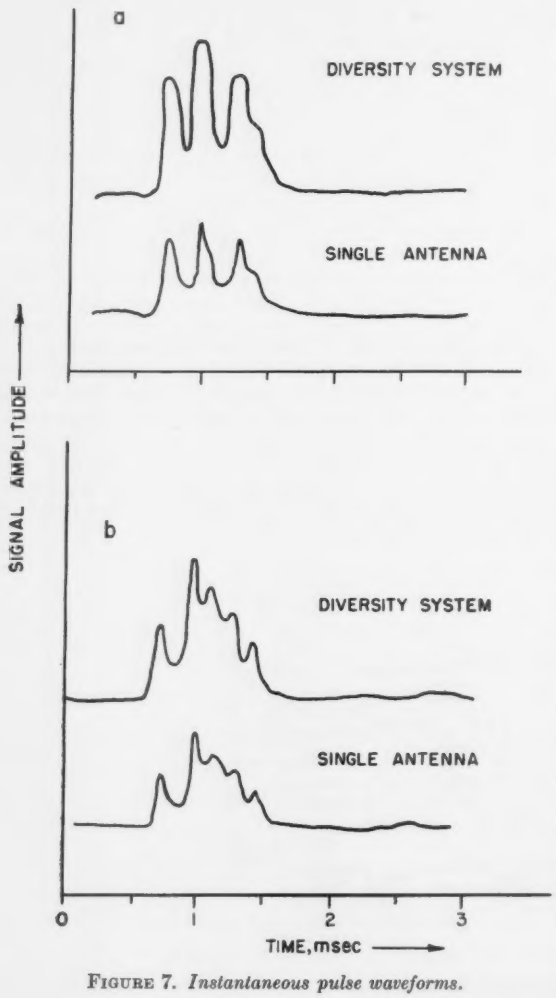


FIGURE 7. *Instantaneous pulse waveforms.*

Using this method of recording, the only attention the apparatus required was occasional adjustment of the crystal frequency, to ensure that the pulse pattern was kept about in the center of the oscilloscope time base.

Generally speaking, the signal was usually detectable from 1000 to about 1530 G.m.t.; on some occasions, the transmissions did not extend over the full period. The signal was usually at its strongest at about 1300 G.m.t., but no detailed measurements of signal strength were made. The type of pattern observed varied considerably with propagation conditions.

Figure 8 shows a simple type of pattern, in which the time base was applied vertically, with delay increasing upward. The sinusoidal variation in delay, with a period of about 10 sec, is due to imperfect synchronization of the pulse. The vertical white dashes are due to pulses of static, of very short duration; the striations in the trace near the left-hand edge are the result of CW interference from a transmitter on an adjacent channel.

In the summer, much stronger echoes are received on days when *Es* ionization is present. As shown in section 2.2, its effect is to generate a group of closely spaced echoes, centered on each order of *F* reflection. Depending on the geometry of reflection, the spacing may be from 0.1 to 0.2 msec. Figure 9(a) is an example of predominantly *Es* propagation, with only one order of *F* reflection present. The four components of the pulse have a mean separation of about 0.15 msec. Figure 9(b) shows an intermediate case, where four *F* echoes are present, with a small amount of *Es* causing each to split into three distinct echoes. In figure 9(c) the *Es* echoes have completely obscured the *F* echo pattern.

4. Comparison With Simple Ray Model

Designating the three pulses in figure 7(a) as *P*, *Q*, and *R*, in order of arrival, the time separation between pulses *P* and *Q* is about 0.24 msec, and between *Q* and *R* is about 0.28 msec. To explain these separations by a constant reflection height would lead to a height of only 200 km, from figure 2: this is clearly incompatible with the ionospheric data of section 2.1.

One possibility for explaining the various modes is in terms of low- and high-angle rays. The curve *GHI* on figure 3 represents a parabolic layer with height of maximum ionization 300 km, semithickness 120 km, and critical frequency 6 Mc/s. The points *G*, *H*, and *I* correspond to low and high angle third order *F* reflections, and a high angle second order reflection, respectively. This configuration produces a set of three modes with very nearly the observed overall separation: however, the two modes *H* and *I* occur with nearly identical delay, rather than the observed .28 msec separation. A departure of the layer from a parabolic form, giving a flattening of the "nose" of the layer, might displace point *I* to *J*, thereby removing the degeneracy. The ray would

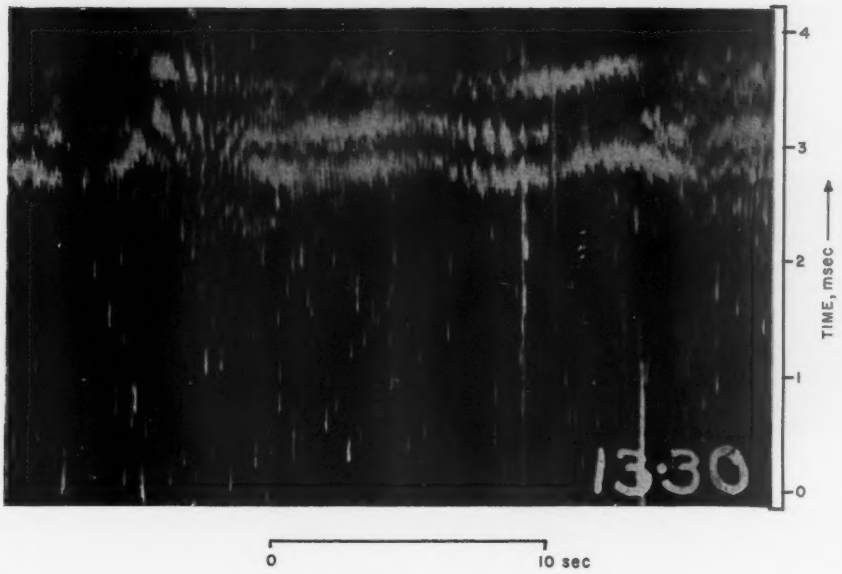


FIGURE 8. *Typical simple pattern.*

then arrive at a very low angle to the horizontal. On the whole, a flattening involving an equivalent vertical incidence group retardation of over 70 km for a frequency increment of less than 0.2 Mc/s seems unlikely, and Kift (1960) has suggested that high-angle modes propagated over great distances are unstable. However, this interpretation cannot be ruled out, particularly in view of the approximations involved in assuming uniform propagation characteristics over the entire path.

An alternative interpretation is that two of the three successive pulses of figure 7(a) have the same order of reflection but are different magneto-ionic components. On a thin-layer hypothesis (fig. 2), if pulses P and R are assumed to be the same component for the third and fourth order reflections, the delay between the two, 0.52 msec, corresponds to a reflection height of about 300 km, and their angles of arrival are about 5° and 11° respectively. If pulse Q is assumed to be a third order reflection, its effective reflection height must be about 335 km in order to account for the additional delay, and its angle of arrival must be about 7°. This might well be accounted for by pulses P and Q being the ordinary and extraordinary third order reflections respectively. Assuming the same parabolic layer as previously, the ordinary and extraordinary ray reflection conditions have been plotted on figure 3 as ABC and DEF respectively. It is suggested that the pulses P, Q, and R may correspond to the points A, D, and B respectively.

Since there appears to be some doubt as to which of these interpretations is correct, the pulses will be referred to by the letters P, Q, and R in subsequent sections.

5. Fading Analysis

Using the analysis of the previous sections as a basis, the principal characteristics of the fading observed over this path will now be described.

5.1. Simple Echo Fading

On a limited number of occasions, only a single order of reflection was present in the pattern. At these times, an easily interpretable fading curve could be obtained by photographic recording; an example is shown in figure 10. The upper and lower traces are the fading curves received at two antennas of a space diversity pair, 600 ft apart in a line perpendicular to the direction of propagation. There is clearly a high correlation between these curves, indicating a distance scale of much greater than 600 ft in the random diffraction pattern.

5.2. Complex Echo Fading

When only a single order of reflection is present, the fading of a pulsed signal is identical with that of a CW signal, and only fluctuations in amplitude are important. If a CW signal is composed of several orders of reflection, however, the possibility of phase interference between the various echoes may give both a faster rate of fading and a smaller structure size on the ground (greater diversity in a space-diversity system).

This comparison may be readily effected by switching the receiving system bandwidth. If a wide bandwidth is used in the receiving system, so that the orders of reflection are resolved, each order will de-

scribe its own fading curve. However, if the bandwidth is narrowed to such an extent that the pulses completely run into one another, thus approximating to the CW case, then the resultant fading will be due not only to the fading of the orders of reflection, but to the effect of phase interference between them.

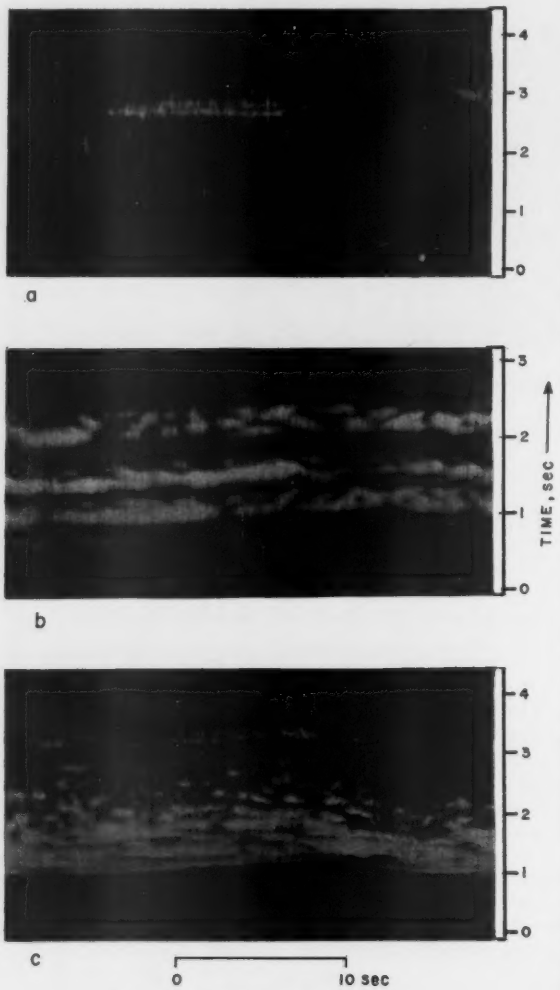


FIGURE 9. *Pattern showing* Es *echoes.*

To illustrate this effect, figure 11 shows four successive stretches of fading record, in which the receiver bandwidth was switched from 8 kc/s to 1 kc/s instantaneously. There is complete diversity between the two 1 kc/s signals in 11*B* and 11*D*, but the 8 kc/s signals of 11*A* and 11*C* display marked correlation. In order to study the rapid fading of the various elements of the pulse waveform in detail, and to relate this to the behaviour of a diversity receiving system, photographs were taken of the pulse waveform at 2-sec intervals. Two examples of these photographs were shown in figure 7. In each case the upper waveform represents the pulse received by a four-fold diversity system, i.e., the mean of four-pulse patterns received on four separate antennas and receivers. The other pulse shown was that received simultaneously on a single antenna.

To remove the possibility of slight random jitter of the pattern, the average of a number of waveforms was first calculated, and is shown in figure 12. With this as a guide, the individual waveforms were synchronized, and peak values read off for the *P*, *Q*, and *R* reflections. The mean values of these quantities are shown with circles on figure 12. They are greater in amplitude than the previously measured peaks of the three pulses, due to slight lack of synchronism in calculating figure 12.

Typical fading curves for the various pulses for a single antenna and for a diversity system are shown in figures 13, 14, and 15.

5.3. Diversity Comparison

Several interesting points are immediately evident from figures 11, 12, and 13. Firstly the depth of fading for the three pulses is quite different, being much less for the *Q* reflection than for either the *P* or the *R*. The best measure of this depth of fading is the ratio of the standard deviation of the fading curve to its mean value. If the depth of fading is defined as *D*, then

$$D=\frac{\sqrt{\overline{x^2}-\overline{x}^2}}{\overline{x}}$$

where x=amplitude of fading curve at any instant. Values of *D* are given in table 2 for the three pulses. For a single aerial, the values of *D* are 0.35 and 0.34

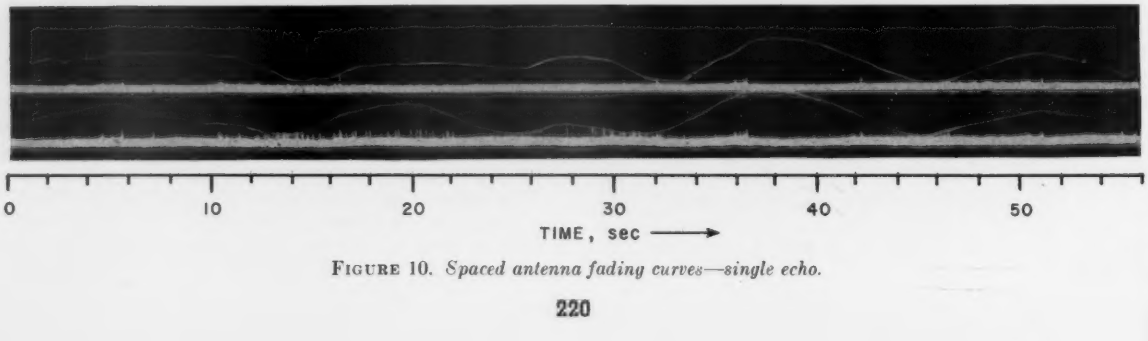


FIGURE 10. *Spaced antenna fading curves—single echo.*

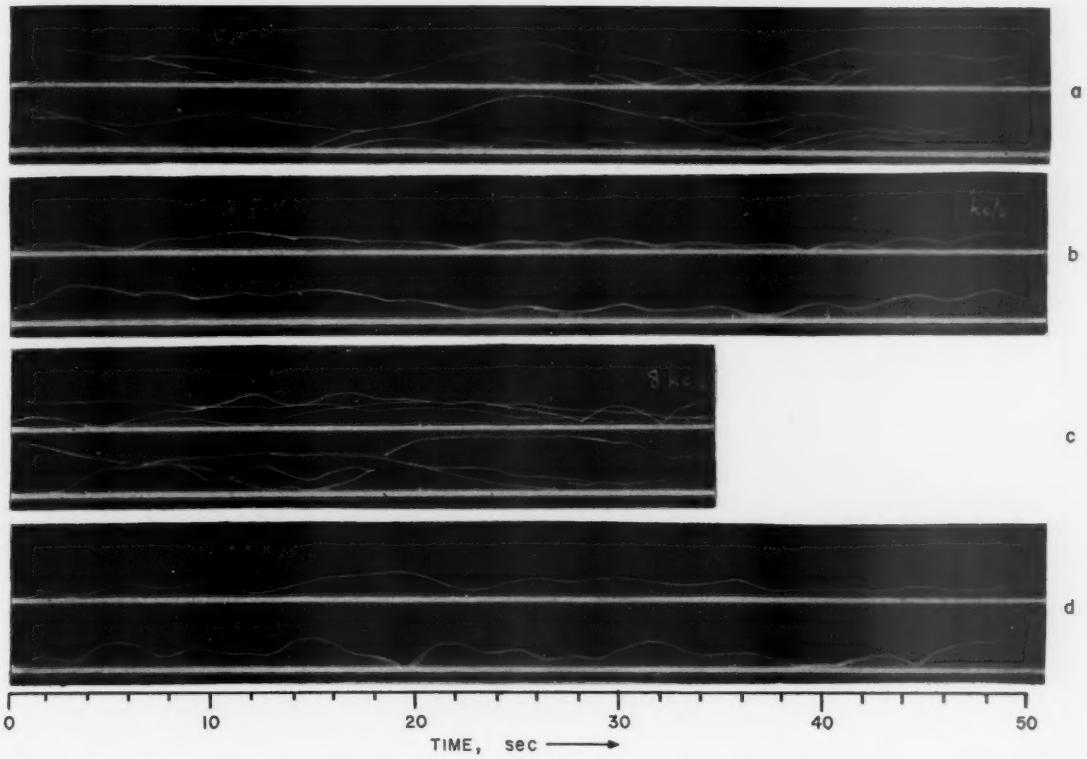


FIGURE 11. *Spaced antenna fading curves—multiple echoes.*

TABLE 2

Pulse	Aerial	Mean signal	D	ρ_1	U
					sec
P	Single	12.7	0.353	0.439	3.0
	Diversity	19.1	.367	.466	2.9
Q	Single	29.1	.174	.450	2.5
	Diversity	43.5	.173	.465	2.9
R	Single	15.5	.343	.343	2.1
	Diversity	24.5	.306	.065	1.5

for the P and R pulses, but only 0.17 for the Q pulse. The greatest value of D which would be expected experimentally would be the value for a signal with a Rayleigh probability distribution, which would give a value of 0.52.

Though the diversity system has virtually no effect on the depth of fading for two of the echoes received, this does not mean that it would also be ineffective for a CW signal; since in CW reception the different orders of reflection may interfere destructively. The only condition for a diversity system to give an improvement in the communication system performance is that the phase relation between the echoes shall be different at the four diversity antennas.

The difference in fading speed of the three pulses should be noted. The fading periods of the P and Q pulses are about 1 cycle in 5 to 10 sec. The fading of the R pulses, however, is very much more rapid. The best quantitative measure of the fading speed is the correlation function of the amplitude. The first ordinate of this is the correlation ρ_1 between the amplitudes of the same pulse in successive snapshots (taken at 2-sec intervals). If it is assumed that the time correlogram of the fading curves has a Gaussian form, i.e.,

$$\rho(\tau)=e^{-6.58\tau^2/U^2}$$

where $\rho(\tau)$=correlation between amplitudes separated by a time interval τ,

U=average time between peaks of fading curve,

="fading period" of curve,

then the fading period of the pulse amplitudes can be calculated. These values of U are also given in table 2.

Another feature of the fading curves is the effect of the diversity system on the fading of the various

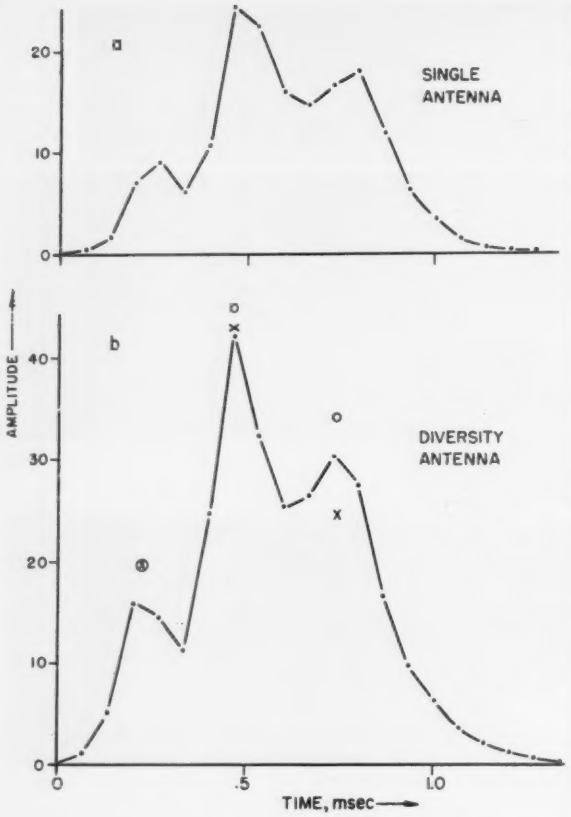


FIGURE 12. *Average pulse waveforms.*

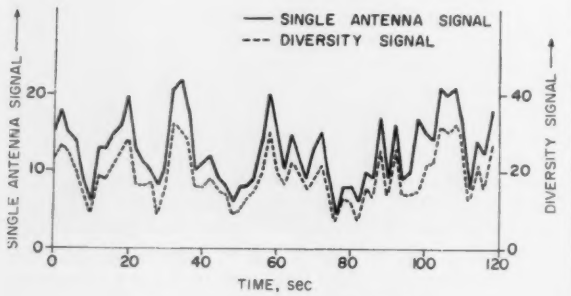


FIGURE 13. P *pulse amplitude.*

pulses. If the amplitudes of the pulses at the four antennas were completely unrelated, then the value of D for the diversity combination of these four should only be half that for the pulses received on a single antenna. In fact, however, as can be seen from table 2, the value of D for the P and Q reflections is nearly the same whether a single aerial or a diversity system is used, indicating that for these pulses the amplitudes received at the diversity antenna were practically identical. Only in the case of the rapidly fading pulse, P, is there an appreciable difference in the values of D.

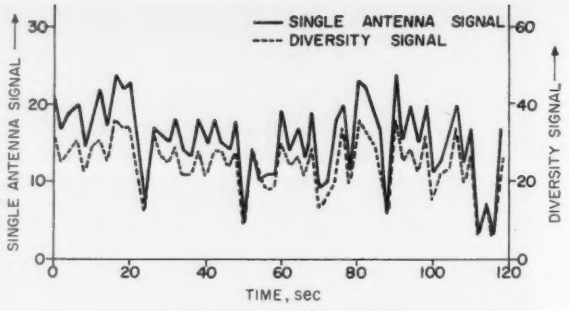


FIGURE 14. Q *pulse amplitude.*

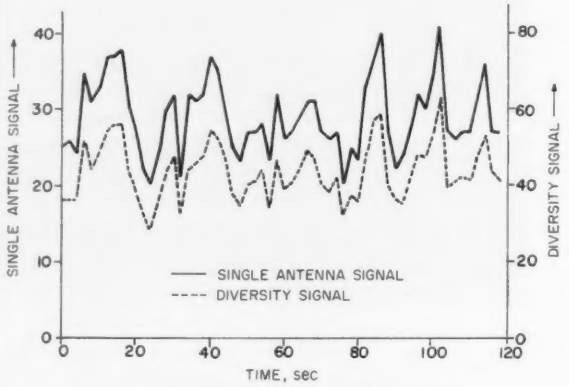


FIGURE 15. R *pulse amplitude.*

A simple relation is derived in the appendix between the depth of fading from a diversity system and the original depth of fading of the single aerial signal, in terms of the average correlation between the signal amplitudes received by the various diversity aerials. The simple theory also correctly predicts the correlation between the single antenna amplitude and the diversity amplitude.

6. Conclusion

The work which has been described was originally intended to clarify the meaning of measurements of diversity at high frequencies. The fading of the different modes of propagation were studied separately, and it was shown that the diversity arose both through diversity of the modes of propagation themselves, and also through phase interference between them; which of these is the major source of diversity depends on the degree of complexity of the pulse pattern.

The need to identify the pulses received led to a consideration of the way in which a steady signal was propagated over long distances; it has been shown that the "sky billiards" approach of geometrical optics gives an adequate explanation of the experimental observations.

It is suggested that when the propagation is simple, i.e., when there is only one predominant order of reflection present, the deversity of a CW signal is due to the irregularities in the reflecting layer; but when the propagation is complex, with many orders of reflection present, the diversity of the CW signal is due principally to phase interference between the orders of reflection.

7. Appendix. Theory of n-Fold Diversity Combination

A diversity signal is formed by averaging n signals $R_1 R_2 \ldots R_n$ to give a signal R_D. It is required to find.

1. The ratio A by which the standard deviation σ of the original signals is reduced by the process of averaging.

2. The correlation of the average with any one of the signals. It will be assumed that the correlation of each signal with every other signal is equal to ρ.

1. $$n^2 \cdot \overline{R_D^2} = \overline{(R_1+R_2+\ldots+R_n)^2}$$

$$= n^2\overline{R}^2 + n\sigma^2[1+(n-1)\rho]$$

$$\therefore A = \sqrt{\frac{1+(n-1)\rho}{n}}$$

$$\text{or } \rho = \frac{nA^2-1}{n-1}.$$

For the R pulse of section 5.3, $A = \frac{0.306}{0.343} = 0.892$

or $\underline{\rho = 0.73}$.

2. $$\overline{R\,R_D} = \frac{1}{n}\Sigma\,\overline{R\,R_D} = \overline{R}^2 + \sigma^2[\rho + (1-\rho)/n]$$

$\therefore$ Correlation between R and R_D

$$= \frac{\sigma^2[\rho+(1-\rho)/n]}{\sqrt{\sigma^2 \cdot \sigma^2[\rho+(1-\rho)/n]}} = A.$$

So, taking the previous example,

correlation between R and $R_D = 0.89$

This compares with the value of 0.887 calculated directly from the observations of section 5.3, indicating the accuracy of this theory.

8. References

Kift, F., The propagation of high-frequency radio waves to long distances, Proc. IEE **107** pt. B, 127–140 (March 1960).

Silberstein, R., A long-distance pulse-propagation experiment on 20.1 megacycles, J. Geophys. Research **63,** 445–466 (Sept. 1958).

Wilkins, A. F., and F. Kift, Characteristics of H.F. signals, Electronic and Radio Engineer **34,** 335–341 (Sept. 1957).

(Paper 65D3–121)

JOURNAL OF RESEARCH of the National Bureau of Standards—D. Radio Propagation
Vol. 65D, No. 3, May–June 1961

Influence of Ionospheric Conditions on the Accuracy of High Frequency Direction Finding[1, 2]

P. J. D. Gething

(August 22, 1960)

The accuracy of fixes obtained by HF direction finding stations has been examined by means of a dispersion factor computed for each fix; this factor is a measure of the consistency of bearings taken from different stations on the same transmission. It is shown that the accuracy is significantly lower during times of ionospheric storms than at times when no storm occurred, and that the effect of the storm is mainly on fixes involving $F2$ layer propagation.

1. Introduction

It is natural to suppose that the accuracy of DF bearings taken on transmissions in the HF band will depend to some extent on prevailing ionospheric conditions. In particular, ionospheric storms might be expected to lead to some deterioration of accuracy. The present paper describes a statistical analysis undertaken to test this hypothesis.

The bearings analyzed were obtained from stations equipped with standard Adcock direction finders. They were taken on both fixed and mobile transmitters, the majority of the transmitters being in the distance range 1,000 to 3,000 km from the DF stations. The true positions of the transmitters were not known in the majority of cases.

The quality of each bearing was estimated by means of the Brooke system. In this system, a variance is assigned to each bearing depending on several factors known to be related to accuracy. The variance components were originally estimated from an analysis of the errors in some thousands of bearings obtained on targets of known position; they are regularly revised on the basis of analysis of current data on such targets. The system is described in detail in an accompanying paper by E. M. L. Beale. It should be noted that the Brooke variances do not take account of variations in ionospheric conditions explicitly, although the components depending on the swing and range of snap bearings are probably correlated with such variations to some extent.

The bearings obtained on a given task from different DF stations do not normally intersect in a point. An electronic computer has been programed to solve the statistical problem of finding the most probable position of the transmitter (Best Point Estimate or BPE) and a surrounding rectangle representing a 90 percent probability region. The Brooke variances are used in the calculation, and affect the position of the BPE and the size and position of the rectangle. The computer also calculates a so-called 'dispersion factor,' the definition of which is given in section 3. This factor is a measure of the consistency of three or more bearings taken from different stations on the same task. The average value of the dispersion factor should be $(n-2)$, where n is the number of bearings. Under adverse conditions the average value would be greater than $(n-2)$.

The average value of dispersion factor has been used throughout the analysis as a measure of accuracy. The advantage of using this factor in a quantitative investigation is that it can be calculated on all tasks without any knowledge of the true positions of the transmitters.

2. Definition of an Ionospheric Storm

Various phenomena are associated with ionospheric storms, the most characteristic being the change in the critical frequency of the $F2$ layer ($fo\ F2$). The critical frequency often shows a brief initial rise, termed the positive phase, followed by a steep fall to a much longer negative phase, during which $fo\ F2$ is below the value expected for calm conditions.

It is conventional to classify storms as moderate or severe. A moderate storm is defined in the B.B.C. monthly propagation reports as one in which $fo\ F2$ lies between 60 percent and 90 percent of the fortnightly running hourly mean $fo\ F2$, and a severe storm as one in which $fo\ F2$ is less than 60 percent of this mean. The B.B.C. reports list days on which severe and moderate storms occurred.

The B.B.C. classification was used in the initial stages of the investigation. Later, in order to determine the magnitude and duration of each storm more precisely, published data from the U.K. vertical sounding stations at Slough and Inverness were examined. The start of the storm was then taken to be the time at which $fo\ F2$, measured at Slough, fell below 90 percent of the monthly median for the corresponding time. The storm became severe when $fo\ F2$ fell below 60 percent of the same median value. The progress of a typical storm is shown in figure 1, where the Slough value of $fo\ F2$ is plotted against time.

[1] Contribution from Government Communications Headquarters, Cheltenham, England.
[2] Paper presented at the Conference on Transmission Problems Related to High-Frequency Direction Finding, at UCLA, June 21-24, 1960.

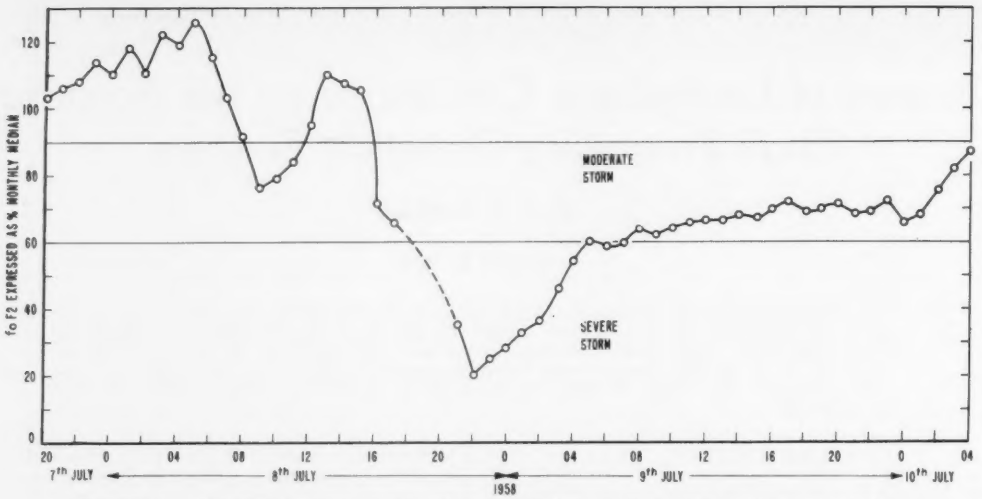


FIGURE 1. *Ionospheric storm on 7th, 8th, 9th, and 10th, of July 1958, readings at slough.*

---------- Indicates discontinuity in readings.

It is, of course, rather arbitrary to define the start and finish of a fairly widespread storm by means of soundings taken at a single station. Comparison of Slough and Inverness data showed that the broad features of the behaviour of *fo F2* were the same at these stations, but that the smaller features of the plotted records were not identical. Measurements from other European sounding stations had not been published when the analysis was undertaken. Since the local times at points of reflection for most of the tasks analyzed did not differ by more than about two hours from the time at Slough, the behaviour of *fo F2* at Slough was probably reasonably representative of conditions over the area concerned.

3. Dispersion Factor

As already explained, the computer calculates a BPE of the position of the target transmitter for each set of bearings fed to it. The observed bearings are then compared with the bearings which would pass through the BPE, and an angular residual formed for each contributing DF station. Each residual is squared and divided by the Brooke variance of the bearing. The dispersion factor is then the sum of the resulting numbers for all the bearings contributing to the fix. The factor is calculated by the computer to one decimal place.

Simple statistical theory suggests that the dispersion factors for n stations should have a χ^2 distribution with $(n-2)$ degrees of freedom. Details of this distribution may be found in most textbooks on statistics, e.g., Yule and Kendall, Introduction to the Theory of Statistics, chapter 20. The mean value of the dispersion factors should be $(n-2)$ and the variance about the mean should be $2(n-2)$.

These results depend on the assumptions usually made in DF analysis, that bearing errors are normally distributed about zero mean and that the errors in bearings taken from different stations on the same task are not correlated. It is known, however, that none of these conditions is strictly satisfied in practice. In particular, it seems that the leptokurtic distributions of bearing errors which occur in practice will lead to a variance rather greater than $2(n-2)$. It is difficult to estimate the effect on the dispersion factors of small systematic errors in bearings and of correlation between bearing errors. These points have not been considered in detail here, since we are concerned more with a comparison of factors under different conditions than with their absolute values.

The dispersion factor will be denoted by Σ_o following the notation used by Beale and others. No convenient name or symbol exists for the quantity $\{\Sigma_0-(n-2)\}$; the symbol Σ_n will be used in this paper. We except from simple theory that Σ_n will be zero on average, positive under adverse conditions and negative under favourable conditions. It will be shown later that, in practice, the average value of Σ_n for a given day tends to be positive.

4. Comparison of Calm and Stormy Days

A period of six months, from July to December 1958 inclusive, was selected for study. According to the B.B.C. reports, severe storms occurred on 15 days in this period and there were 49 days with moderate storms.

It was decided to compare values of Σ_n on days of severe storms with values on samples of calm days. Saturdays and Sundays were not used in either sample because computer results for tasks on these days had not been calculated. There then remained 12 days of severe storms for study. Four calm days were chosen, one each in July, September, October, and November, thus giving a reasonable distribution throughout the period. Days on which magnetic storms or Dellinger fades occurred were not selected for this sample, but otherwise the days were chosen at random.

The results are set out in table 1 below. The average value of Σ_n for all stormy days is +1.27 and is significantly greater than the value of +0.57 on calm days.

It is clear from table 1 that propagation conditions were worse from the point of view of DF accuracy on days of severe storms than on calm days, as expected. The question next arises as to whether this effect applies to all frequencies, or is more marked in a particular part of the HF spectrum. Table 2 shows the value of Σ_n for 1 Mc/s intervals for the same sample of calm and stormy days. The difference between the results is more marked on the first and second frequency bands (i.e., up to 10 Mc/s) than on the third band, but all are affected to some extent.

In order to investigate further the effect of storms on accuracy, it was decided to use Slough measurements of *fo F2* to define periods of moderate and severe storms as described earlier. The period considered was extended to June 1959. Table 3 shows values of Σ_n during the moderate and severe phases of the storms considered.

TABLE 1. *Comparison of calm days with days of severe storms*

Calm days		Stormy days	
1958	Average value of Σ_n	1958	Average value of Σ_n
14th July	+0.5	8th July	+1.2
22d September	+0.6	9th July	+1.2
14th October	+0.7	18th August	+1.2
14th November	+0.5	4th September	+1.4
		5th September	+1.1
All 4 days	+0.57	23d October	+1.4
		24th October	+1.6
		28th October	+1.4
		4th December	+1.0
		5th December	+1.3
		17th December	+1.2
		18th December	+1.3
		All 12 days	+1.27

TABLE 2. *Dependence of Σ_n on frequency*

Frequency	Calm days	Stormy days
	Average value of Σ_n	Average value of Σ_n
Mc/s		
1-2		+2.7
2-3	+0.3	+1.9
3-4	+1.5	+2.0
4-5	+1.2	+1.6
5-6	+0.2	+1.6
1-6	+.95	+1.72
6-7	+.8	+1.5
7-8	+.4	+1.4
8-9	+.2	+1.3
9-10	.0	+1.4
6-10	+.47	+1.41
10-11	−.3	+0.2
11-12	.0	+1.0
12-13	+.3	0.0
13-14	+.2	+.4
14-15	−.6	−.1
>15	+1.0	+.9
>10	+0.17	+.48

TABLE 3. *Σ_n for moderate and severe phases of storms*

1958	Average value of Σ_n	
	Moderate phase	Severe phase
18th August	+1.1	+1.4
4/5th September	+1.2	+1.4
23/24th October	+1.5	
28th October	+1.4	
4/5th December	+1.4	
17/18th December	+2.1	
1959		
9th January	+1.0	+2.9
28th February/1st March	+0.9	
1/2d March		+1.8
2/3/4th March	+0.7	
23/24/25th April	+1.2	
Combined result	+1.17	+1.66

The storm on 8th and 9th July 1958 had to be omitted because the relevant logs had been destroyed.

The average value of Σ_n shows the expected increase under increasingly stormy conditions.

It should be noted that several dispersion factors were affected by the rejection of bearings showing large departures from the BPE. Either machine rejection or operator rejection may be applied to the results calculated by the computer; the criteria for rejection will not be described in detail here, but depend in both cases on an examination of the dispersion factor. Rejection of a bearing which is obviously "wild" normally leads to a considerable reduction in this factor.

An examination of computer results showed that a higher percentage of bearings were rejected on stormy days than on calm days. On days of severe storms about 3.9 percent of all bearings were rejected, whereas on the calm days examined only 1.2 percent were rejected. This result is to be expected in view of the increased size of the dispersion factors on stormy days. The greater percentage of rejections on stormy days tends to reduce the difference between the average value of Σ_n on calm and stormy days; hence the difference would be even more marked than in table 1 if the same percentage of bearings had been rejected on both types of day.

A sort was carried out to decide whether bearings taken by any particular DF operators gave rise on average to abnormally large or small contributions to the dispersion factors. Results were inconclusive.

5. Dependence on Mode of Propagation

Individual tasks on two calm and two stormy days were examined with a view to determining the probable mode of propagation.

The range from the BPE for the transmitter to London was taken to be a reasonable approximation to the true distance of the transmitter from the DF station. The maximum usable frequency (MUF) for the same path was derived for the *E*, *F*1, and *F*2 layers from the U.S. Central Radio Propagation Laboratory (CRPL) predictions. These frequencies were then compared with the known signal fre-

580999—61——3

quency, and possible modes of propagation noted. By suitable division of the path it was possible to decide whether multiple-hop modes were likely.

The tasks were then sorted into two groups depending on whether or not propagation was via the $F2$ layer. The sort was necessarily approximate, since neither the predicted frequencies nor the distances used were free from error. However, the assignment of mode of propagation was expected to be accurate enough to reveal any large difference in DF accuracy dependent on the layer involved.

Results of the sort are shown in table 4. Σ_n was again used as a measure of the scatter of the bearings. From the table it appears that direction finding on signals propagated via the E or $F1$ layers is less affected by ionospheric storms than direction finding on signals propagated via the $F2$ layer.

TABLE 4. *Dependence of Σ_n on probable mode of propagation*

	Average values of Σ_n	
	Two calm days	Two stormy days
E and F1	+0.7	+0.5
F2	+ .6	+1.5

6. Effect of Storms on Systematic Errors

The errors of bearings taken on targets of known position were calculated and averaged for each DF station for three calm days and three stormy days. On the calm days, the average systematic error was $+0.6°$; on the stormy days it was $+0.1°$.

The nature of systematic errors is not completely understood, and it is therefore not possible to draw any conclusions from the reduction when conditions are stormy. Each systematic error probably represents a combination of effects due to instrumental errors, site errors not removed by calibration and ionospheric effects.

7. Conclusions

The investigation has established that DF accuracy deteriorates during ionospheric storms and is worst during severe storms. The effect of a storm may extend over the whole HF band, but is more marked below 10 Mc/s than above. The main effect seems to be on signals propagated via the $F2$ layer. Systematic errors during storms generally become less positive.

The results imply that the Brooke variances should be increased during storm periods; such an increase would lead to rather larger probability rectangles on fixes taken in these periods.

The present investigation demonstrates the advantages of adopting an objective classification system for bearings and of using an electronic computer to reduce DF data on a uniform basis and in a form suitable for analysis.

The author thanks Miss A. Partridge and Mr. R. M. Rampling who carried out most of the analysis and who made many valuable suggestions.

This paper is published by permission of the Director, Government Communications Headquarters, Cheltenham, England.

(Paper 65D3–122)

JOURNAL OF RESEARCH of the National Bureau of Standards—D. Radio Propagation
Vol. 65D, No. 3, May–June 1961

Phase Difference Observations at Spaced Aerials and Their Application to Direction Finding[1, 2]

W. C. Bain

(August 22, 1960; revised December 15, 1960)

The measurement of phase differences at two aerial pairs, each spaced by three to four wavelengths, was used to give mean bearings over periods of three to five minutes on certain transmitters with frequencies near 6 Mc/s. The results showed that site errors had probably been reduced by this system to less than 0.5 deg, and possibly to 0.2 deg, which is a value derived on theoretical grounds. The variance of the mean bearings due to lateral deviation was found to be slightly greater than anticipated. However, the results were such as to indicate that the performance of a wide-aperture direction finder should not fall seriously below that which would be expected of it theoretically on the basis of existing knowledge.

1. Introduction

In the winter of 1954–55 a number of observations were carried out at the Radio Research Station in which phase differences were measured between the signals picked up at aerials spaced by several wavelengths. Recently it was noted that these results were relevant to the problem of determining the accuracy obtainable from a wide-aperture direction finder, an instrument capable of improved performance in conditions of multimode propagation [Bain, 1956]. It therefore seemed desirable to present this account of the work and its implications here.

The four aerials used were short unipoles; they were placed on the circumference of a circle of diameter 180 m so that one pair of aerials was alined in the north-south direction and another in the east-west direction. Inside a hut at the center of the circle each of these pairs of aerials was connected to phase-measuring equipment of the type described by Ross, Bramley, and Ashwell [1951]. When the receivers were tuned to the desired station, the overall phase balance of the system was checked, and the phase differences between the signals picked up on the north-south and east-west pairs were recorded photographically at three-second intervals for a period not exceeding five minutes, thus giving about 70 values for each. Ionospheric conditions were characteristic of the time in question and were never greatly disturbed. Observations were confined to the hours of daylight.

It is important to note that such a system is not directly suitable for use as an operational direction finder. Not only does it have bearing ambiguities, but the bearing calculated from the phase differences is not in general correct when two rays are incident on it from the same azimuth.

[1] Contribution from Department of Scientific and Industrial Research, Radio Research Station, Ditton Park, Slough, Bucks., England.
[2] Paper presented at the Conference on Transmission Problems Related to High-Frequency Direction Finding, at UCLA, June 21–24, 1960.

2. Results

2.1. Method of Analysis

In reducing the observations a mean $\bar{\psi}$ of each distribution was calculated in the following way. If there are n phase differences ψ_k in each distribution, then

$$\bar{\psi} = \arg\left\{\sum_{k=1}^{n} \exp\,(j\psi_k)\right\}.$$

This avoids the ambiguities which can arise if an ordinary arithmetic mean is used.

The standard deviation (σ) of the distribution was taken to be the root-mean-square departure from $\bar{\psi}$. For a completely uniform distribution of phase differences the standard deviation would have the value 104° approximately. A finite number of random phase differences will, of course, give a standard deviation somewhat less than this, probably about 94° for the numbers of observations used here.

2.2. Observed Values

The observations were concentrated upon a few stations which had been found from a previous experiment to give widely spread phase difference distributions. Of these, the most frequently used were the following. The distances and bearings shown are from the site of the equipment at Winkfield, near Windsor (Berkshire).

It will be seen that with the aerial spacing used here, the diameter of the system will be in the range of three to four wavelengths for reception of these stations.

It was indeed found that very large standard deviations were occasionally obtained; a few were so great that they could have arisen from a random distribution of phase differences. A complete set of

the standard deviations obtained on these three stations is shown in figure 1. These can be seen to be well spread out over the range of possible values. However, the north-south and east-west standard deviations are correlated to some extent, for if one is large the other also tends to be large. In the case of Arganda, the north-south phase differences were much more widely spread than the east-west; the mean standard deviations were in the ratio of 2.0 to 1. This probably happened because the Arganda bearing is only 10° away from the north-south direction, combined with the fact that the incident radiation was spread much more in elevation than in bearing.

3. Bearings Deduced From Phase Measurements

If the phase difference between north and south aerials is ϕ_1, and that between east and west aerials is ϕ_2, then the bearing γ can be taken to be derived from the formula

$$\tan \gamma = \phi_2/\phi_1.$$

With widely spaced aerials the values of ϕ_1 and ϕ_2 obtained from the measurements may differ from the true values by some integral multiple of 2π. However, as the position of the transmitter was known in these experiments the ambiguities could be eliminated by choosing the multiples of 2π so as to bring the bearing as nearly as possible to the true bearing of the transmitter. The multiple of 2π which had to be used never exceeded 3 in any of the present instances.

Mean bearings were calculated by using the mean phase differences $\bar{\psi}$ from each of the groups of observations; the results obtained are summarized in table 2, which gives results for the mean of all mean bearings for each station.

It can be seen that the results are consistent with the hypothesis that there is no systematic error in the observations. However, it is useful to consider in some detail the possible errors which might arise. They can be divided into three main categories for the present purpose:

1. Rapidly fluctuating errors, with periods of the order of seconds; they may be due to wave interference or polarization errors.

2. Lateral deviation errors, with periods of the order of tens of minutes, due to the effect of ionospheric tilts.

3. Site errors, which will be considered here to be virtually constant with time.

The first two of these errors would be expected to give rise to a spread of values in the mean bearings obtained for each transmitting station, but the third would lead to a systematic error. On the assumption that the site error due to a narrow-aperture Adcock direction finder is 2.5°, it can be estimated from the curves for site-error reduction given by Hopkins and Horner [1949] that the site error for each of the stations in this wide-aperture direction finder would be of the order of 0.2°. Now the range in the fiducial limits for the mean as given in table 2 is sufficiently great to conceal altogether an effect of this magnitude. It is still, of course, possible that the site error of the system is indeed as small as this, but the most that can be said from the experimental evidence here is that it is probably less than 0.5°. This is certainly a worthwhile improvement on a narrow-aperture direction finder. It should be borne in mind that this result was achieved despite the large fluctuations in phase difference commonly found with the transmitters listed in table 1.

The magnitude of the spread of the mean bearings can also be examined to see if it is in accord with what would be expected. To do this some estimate must be obtained of the size of the errors due to rapid fluctuations and to lateral deviation. The error variance (σ_M^2) due to rapid fluctuations can be assessed from a knowledge of the variance of the north-south and east-west phase differences (σ_1^2 and

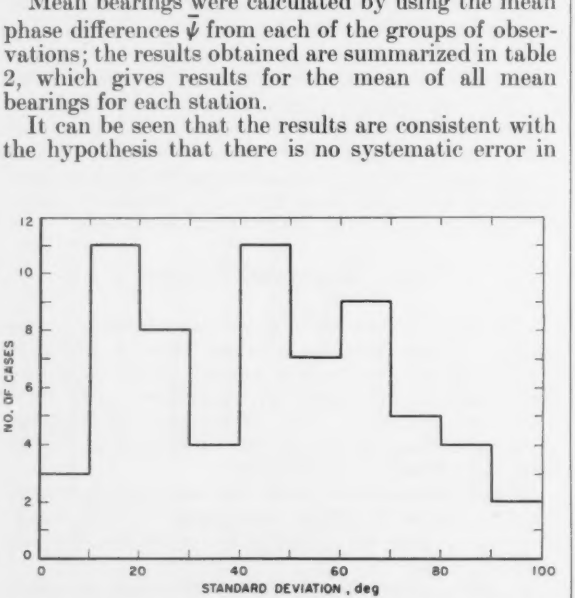


FIGURE 1. *Histogram of the number of cases of standard deviation of the phase difference distributions with values as shown.*

TABLE 1. *Details of stations observed*

Location	Frequency	Distance	Bearing
	Mc/s	*km*	*deg*
Allouis, France	6.20	520	155.0
Arganda, Spain	5.99	1270	191.0
Schwarzenburg, Switzerland	6.16	780	128.1

TABLE 2. *Errors in the consolidated mean bearing*

Station	No. of groups of observations	Bearing error consolidated mean	90% fiducial limits for consolidated mean
		deg	*deg*
Allouis	13	−0.1	−0.8 to +0.6
Arganda	5	−.4	−.8 to .0
Schwarzenburg	13	+.2	−.3 to +.7

σ_2^2, say) in each group. If there are n_1 north-south phase differences and n_2 east-west phase differences, and θ is the bearing of the transmitter, then a reasonable estimate of σ_M^2 is given, as is shown in the appendix, by

$$\sigma_M^2=\left(\frac{\cos\theta}{2\beta}\right)^2\frac{\sigma_2^2}{n_2}+\left(\frac{\sin\theta}{2\beta}\right)^2\frac{\sigma_1^2}{n_1},$$

where

$$\beta=\frac{\pi d}{\lambda}\cos\delta,$$

d=the diameter of the system,
λ=the wavelength, and
δ=the angle of elevation of the incident wave.

Data about the magnitude of ionospheric tilts in the F region can be used to give the variance due to lateral deviation. The figure of 1.2° determined by Bramley [1953] has been adopted here as the standard deviation of the magnitude of ionospheric tilts in one plane. The expected lateral deviation variance for each station was found from this, again with the assumption that the reflection was from the $F2$ layer.

In table 3 the variances of the individual mean bearings as obtained in two different ways are compared; if the estimates made are correct the figures in the last two columns should not be significantly different.

The number of observations on Arganda is too small to give much hope of establishing a significant difference between variances, and in fact the values shown in the last two columns for this station are not significantly different. However, for both Allouis and Schwarzenburg it looks as if the estimated variance may be too small, for in each case the difference between observed and estimated variances is nearly significant at the 95 percent level. As the rapid-fluctuation variance has been estimated from internal evidence, it looks as if this difference is associated with the lateral-deviation component. However, for present purposes it is sufficient to note that the observed values are not very far from the estimated ones, and there is nothing about these results to suggest that our ideas are seriously wrong about how to apply known information about the ionosphere and direction finding to predict the performance of a wide-aperture direction finder.

TABLE 3. *Estimated and observed variances*

Transmitter	Estimated variance due to rapid fluctuations, σ_M^2	Estimated variance due to lateral deviation	Estimated variance of individual mean bearings (sum of previous 2 columns)	Observed variance of individual mean bearings
	deg^2	deg^2	deg^2	deg^2
Allouis	0.08	1.17	1.25	2.05
Arganda	.10	0.32	0.42	0.14
Schwarzenburg	.07	.54	.61	1.08

4. Conclusions

The mean bearing indicated by the phase-difference measuring system considered here, which is of three to four wavelengths in aperture, is likely to be accurate to ±0.5°, and may be as good as 0.2°. To achieve this accuracy time-averaging over a few minutes had to be applied, but in an operational wide-aperture direction finder the fast bearing fluctuations should be smaller than was the case here and should require less time averaging to reduce them to an acceptable level. In the practical direction finder the time required to reduce the variance due to lateral deviation of the rays will probably be the same as here, and several hours would be needed to average out its effects. In the present experiments the variance ascribed to lateral deviation turned out to be somewhat larger than expected, but the difference was not so great as to suggest that large unanticipated effects were present. There are, therefore, grounds for hope that wide-aperture direction finders will be capable of the performance expected of them on the basis of existing knowledge.

This work was carried out as part of the program of the Radio Research Board of the Department of Scientific and Industrial Research, and this note is published by permission of the Director of Radio Research.

5. References

Bain, W. C., The theoretical design of direction-finding systems for high frequencies, Proc. I.E.E. **103,** pt. B, p. 113 (1956).

Bramley, E. N., Direction-finding studies of large-scale ionospheric irregularities, Proc. Roy. Soc. **A220,** p. 39 (1953).

Hopkins, H. G., and F. Horner, Direction-finding site errors at very high frequencies, Proc. I.E.E. **96,** pt. III, p. 321 (1949).

Ross, W., E. N. Bramley, and G. E. Ashwell, A phase-comparison method of measuring the direction of arrival of ionospheric radio waves, Proc. I.E.E. **98,** pt. III, p. 294 (1951).

6. Appendix. Error Variance in the Mean Bearing of a Group of Observations Due to Rapid Fluctuations in the Phase Readings

The true bearing here will be taken to be given by the expression

$$\gamma=\text{arc tan}\,(\phi_2/\phi_1),$$

where $\phi_1=2\beta\cos\gamma$, $\phi_2=2\beta\sin\gamma$.

The error in bearing $\delta\gamma$ is then given by

$$\delta\gamma=\frac{\cos\gamma\delta\phi_2-\sin\gamma\delta\phi_1}{2\beta}.$$

The north-south phase difference, which is here $(\phi_1+\delta\phi_1)$, has a variance σ_1^2 given by

$$\sigma_1^2=\overline{(\delta\phi_1)^2},$$

using the bar to denote an average. If there are n_1 observations of this phase difference, the variance of the mean is given by $\overline{(\delta\phi_1)^2}/n_1$. Similarly, the variance of the mean of the east-west phase distribution is $\overline{(\delta\phi_2)^2}/n_2$. The mean value of the bearing error is given by

$$\overline{\delta\gamma}=\frac{\cos\gamma\overline{\delta\phi_2}-\sin\gamma\overline{\delta\phi_1}}{2\beta}.$$

Hence the variance of the mean bearing is given by

$$\sigma_M^2=\left(\frac{\cos\gamma}{2\beta}\right)^2\frac{\sigma_2^2}{n_2}+\left(\frac{\sin\gamma}{2\beta}\right)^2\frac{\sigma_1^2}{n_1}.$$

This derivation applies only when $\delta\phi_1$ and $\delta\phi_2$ are uncorrelated. However the formula has been checked by calculating bearings for each individual pair of phase differences for certain groups; the standard deviation obtained has been in each case quite close to the value given by the above formula.

(Paper 65D3–123)

JOURNAL OF RESEARCH of the National Bureau of Standards—D. Radio Propagation
Vol. 65D, No. 3, May–June 1961

Research at the National Bureau of Standards Applicable to Long-Distance Location and Direction-Finding Problems[1, 2]

R. Silberstein

(September 8, 1960; revised December 30, 1960)

Research pertaining to radio-location and direction-finding problems conducted at the National Bureau of Standards since 1941 includes evolution of a technique for determining polarization error, study of non-great-circle bearings, the development of a rapid-scanning directional antenna, and the development of Loran-C with its precision timing capabilities.

1. Introduction

Work in long-distance radio-location and direction finding at the National Bureau of Standards began in the Radio Section in Washington, D.C., which became the Interservice Radio Propagation Laboratory (IRPL) during World War II and later became the Central Radio Propagation Laboratory (CRPL) now located at Boulder, Colorado.

2. Polarization Error Studies During World War II

A direction finder may be responsive to vertical or horizontal polarization or to both polarizations phased in various ways, but one polarization may produce an undesired response which causes an error in the reading. This undesired response is a function of the geometrical properties of the instrument, its mountings and various accessories, such as RF cables. A combination of modes in the downcoming ionospheric wave results in more or less random polarization. From instant to instant the relative amplitudes and phases of the wanted and unwanted components of the arriving wave keep changing causing the polarization error to change.

It would be a very difficult thing to compare direction finders and evaluate their polarization errors by observing sky waves, for which bearings depart from their true great circle values for a variety of causes. Some simple test technique was desired which would offer an estimate of the polarization error propensities of any particular direction-finder design. The first test of this type had been evolved by Barfield in England. A balloon with a target transmitter and antenna producing a circularly-polarized wave at a downcoming angle of 45° was placed near the equipment under test. The polarization error observed was called the "standard wave error."

At NBS, a group under the late Harry Diamond devised techniques whereby pickup factors for the desired and undesired polarizations were determined from fixed target transmitters. These pickup factors were then treated analytically in accordance with a procedure designed for each type of direction finder and polarization errors determined for any set of conditions.

In the case of the simple Adcock antenna, the desired polarization is parallel to the plane of propagation (vertical when the downcoming angle is zero) and the undesired polarization is perpendicular to the plane of propagation (horizontal).

The pickup factor for a wave polarized parallel to the plane of propagation was called h. That for a wave polarized perpendicularly to the plane of propagation was called k. If h and k were known then the polarization error could be written (in the notation used by the Diamond group) as:

$$\tan \epsilon = \frac{k}{h} \frac{E_{\perp}}{E_{\parallel z} \cos \psi}$$

where

ϵ is the polarization error angle,
$E_{\perp}$ is the resultant total field vector perpendicular to the plane of propagation (i.e., horizontal polarization) after combination of the incident and ground reflected wave,
$E_{\parallel z}$ is the resultant total field vector normal to the ground (vertically polarized) after the combination of the ground wave and ground-reflected wave polarized parallel to the plane of propagation
ψ downcoming angle

A figure of merit was proposed for the case of equal resultant fields with the direction finder on the ground. Here $\tan \epsilon_o = (k/h)$, either k/h or ϵ_o being the figure of merit.

The advantage of the method was that h and k could be measured on the ground. One type of error called "radiator parallax" occurred in measuring k in direction finders like the balanced H Adcock.

[1] Contribution from Central Radio Propagation Laboratory, National Bureau of Standards, Boulder, Colo.
[2] Paper presented at the Conference on Transmission Problems Related to High-Frequency Direction Finding, at UCLA, June 21–24, 1960.

This was due to a net unwanted vertical pickup in the system when a horizontal dipole was used as a source of perpendicular polarization. In the case of spaced vertical loops another type of error called "collector parallax" was studied. The forward tilt component of the target transmitter's magnetic vector produced an unbalanced voltage in the system. Each type of error could be minimized by increasing the distance to the target transmitter.

The tests for determining h and k were actually conducted with both the direction finder and the target transmitter on wooden platforms above ground. The elevated platforms were especially useful in obtaining a strong horizontally-polarized ground wave for measuring k. The importance of this arrangement is attested to by the author, who at one time attempted to measure the horizontal field strength over a stretch of swamp where vertical polarization was attenuated about 70 db less than horizontal. The transmitting antenna was a wire dipole. Minor unbalances of this antenna as it swayed in the breeze produced great surges of output in the field strength meter as incidental pickup of the vertical polarization responded to the vertical component of the unbalanced antenna system.

3. Radio Propagation Effects on High Frequency Direction Finders (World War II Period)

In this project several receiving stations took routine bearings on distant stations. Results were compared with predicted path MUF's and field strengths. Much attention was paid to an old phenomenon which was noted by the British in the 1930's [Keen, 1947] and which has recently been attracting attention again, namely, propagation of signals by a quasi-scattered non-great-circle mode after failure of the regular path MUF.

The British results were duplicated again in various ways during World War II. For instance, when a BBC broadcast beamed at North Africa was monitored in Virginia, the bearings taken during the hours of predicted MUF above the operating frequency indicated that the bulk of the received energy was arriving via the great circle path. After predicted MUF failure of the eastern control point, however, the signals exhibited reduced fluctuating field strength with widely swinging bearings suggesting a scattering region in Africa. It was assumed that the MUF was still high enough to the south of England for propagation to the scattering region from which energy was being received in Virginia by regular propagation over a more southerly path. In those days it was assumed, following T. L. Eckersley, that high-frequency backscattering and sidescattering occurred in the E region rather than at the ground where most of it is now known to take place, but the general assumption of propagation from a scattering region was valid.

Frequently European stations in the lower part of the high-frequency band, from which great-circle propagation via the regular layers was not possible, were observed to have a northerly bearing deviation when field strength was weak, suggesting a mode involving an Auroral Zone gradient. These studies are called to mind by the results of recent Canadian experiments on 40 Mc/s BBC television in which both types of bearing deviations were seen.

Similar observations of signals from United States stations were readily explained, but a group of widely deviated bearings of a Singapore station, observed at Sterling, Va., were never fully understood.

4. Recent Developments in Techniques

A precise system of direction finding, developed for use on spherics at VLF, has potentialities for other uses [Hefley, Linfield, and Davis, 1961]. The name "Ephi" was chosen for the system because the bearing of the transient signal is determined from the relative phase (phi) of the vertical electric field (E) received at spaced antennas. In the system three vertical antennas are erected at the corners of an equilateral triangle of sides 1/5 to 1/10 wavelength. A transient signal from a given direction arrives at each antenna at a different time. These arrival times are not compared directly; instead, sets of fixed time delays are inserted in each channel, each set defining an azimuthal sector. Whenever a spheric occurs in such a sector the pulses arrive nearly simultaneously at the outputs of their respective channels and trigger a coincidence circuit. Thus the number of spherics in any sector may be counted. The system is superior to the conventional crossed-loop system in that it works on only vertical polarization. Also it lends itself to automatic readout much more readily than a system employing oscilloscopes.

Testing of a new type of wide-aperture direction finder using electronic scanning has just been completed at CRPL [Cottony and Wilson, 1961]. The model tested had a beamwidth of 6½° and scanned over an arc of 45°. Although the present operating frequency is near 40 Mc/s special application of the basic features to high frequency work appears to be feasible.

A theoretical paper was written on a proposed experimental technique for measuring the angle of arrival, azimuth and polarization of a downcoming wave by the use of a combined 4-element Adcock and crossed-loop direction finder [Wait, 1959].

5. Loran C as a Source of Precise Timing

A system of Loran stations on 100 kc/s, known as the East Coast chain has recently been installed with a master station at Cape Fear, N.C., and two slave stations at Martha's Vineyard, Mass., and Jupiter Inlet, Fla. [Doherty, Hefley, and Linfield, 1961]. These stations employ a recently declassified system of navigation formerly known as Cytac.

In Loran C the difference between the differentiated signal envelope and the signal itself yields a wave crossing the zero axis at a point so well-defined that it can be used to eliminate ambiguity of a single cycle

in timing a point on the wave. Synchronization is accomplished entirely by computer circuits.

A slave station can be held to 0.1 μsec or better. The timing accuracy using the Loran C ground-wave emissions is 1 μsec or better over land for ranges to about 1,500 miles and 0.1 μsec over sea for ranges to about 2,000 miles. Much greater distances can be covered by sky-wave synchronization with errors of only a few microseconds. Because of the sporadic impulsive nature of atmospheric noise and the fact that the Loran C pulses arrive at a 20-cycle per second rate, it is possible to synchronize oscillators with a 40 db ratio of average noise power to signal power.

The advantage of a positioning system with an accuracy 1,000 times greater than that for independent clocks compared with WWV sky-wave signals (assuming a 1-min integration time) is obvious.

6. References

Cottony, H. V., and A. C. Wilson, A high-resolution rapid-scan antenna, to be published in J. Research NBS **65D** (Radio Prop.) (1961).

Doherty, R., G. Hefley, and R. F. Linfield, Timing potentials of Loran C, to be published in Proc. IRE 1961.

Hefley, G., R. F. Linfield, and T. L. Davis, The Ephi system for VLF direction finding, J. Research NBS **65C** (Eng. and Instr.) No. 1, 43 (Jan.-March 1961).

Keen, R., Wireless direction finding, pp. 410–415 (Iliffe & Sons, 4th ed., 1947).

Wait, J. R., Downcoming radio waves, measurement of characteristics, Electronic & Radio Engineer 36, No. 3, pp. 106–107 (March 1959).

(Paper 65D3–124)

JOURNAL OF RESEARCH of the National Bureau of Standards—D. Radio Propagation
Vol. 65D, No. 3, May–June 1961

Design for Spinning Goniometer Automatic Direction Finding [1, 2, 3]

W. J. Lindsay and D. S. Heim

(September 29, 1960)

This paper discusses some aspects of instrumentation design in making a spinning goniometer radio direction finder essentially automatic in operation. Also discussed is the application of narrow band synchronous post detection filtering for improving the bearing sensitivity.

1. Introduction

In this paper various aspects of the problem of automating a spinning goniometer radio direction finder (hereafter abbreviated SGDF) are discussed. These considerations involve the handling of the directional information and do not concern other control operations such as tuning, switching, etc. Such a system is desirable for use as a research tool and to study the feasibility of applying various concepts operationally. Although the work was done on a AN/TRD–4 unit, the concepts may be applied to any SGDF. The availability of automatic operation does not preclude the operation of the equipment either partially or entirely in the usual manual fashion.

In a narrow aperture SGDF the simulation of a spun antenna is achieved by balanced amplitude modulation of the incoming wave at a rate set by the spin frequency of the goniometer. If sense voltage (from an omnidirectional antenna) is added to this, the effect is to restore the carrier component and we have a conventional amplitude modulated signal. These two modes of operation will be referred to as the DSB and the AM cases respectively. The AM wave may be expressed as:

$$e(t)=E(1+k\cos\omega_1 t)e^{j\omega_0 t} \tag{1}$$

Taking the real part:

$$Re[e(t)]=E\cos\omega_0 t+\frac{KE}{2}[\cos(\omega_0+\omega_1)t+\cos(\omega_0-\omega_1)t] \tag{2}$$

The detected envelope of the AM wave is;

$$e_d=[-(1+k\cos\omega_1 t)] \tag{3}$$

While the detected envelope of the DSB wave (AM wave minus the carrier) is:

$$e_d=\frac{-K_1}{\pi}\left[1-\frac{2}{1\cdot 3}\cos 2\omega_1 t-\frac{2}{3.5}\cos 4\omega_1 t-\ldots\right] \tag{4}$$

In the above expressions ω_1 is the modulation frequency of the spinning goniometer while ω_0 is the center frequency of the incoming radio frequency wave. It will be noted that in (2) the sideband frequencies are displaced from center frequency only by the spin rate of the goniometer, which in the present case is 30 c/s; also the fact that in the detected envelope of the DSB wave only even harmonies of the modulation frequency appear, there is no fundamental component. The detected envelopes of both the DSB and the AM waves are shown in figure 1.

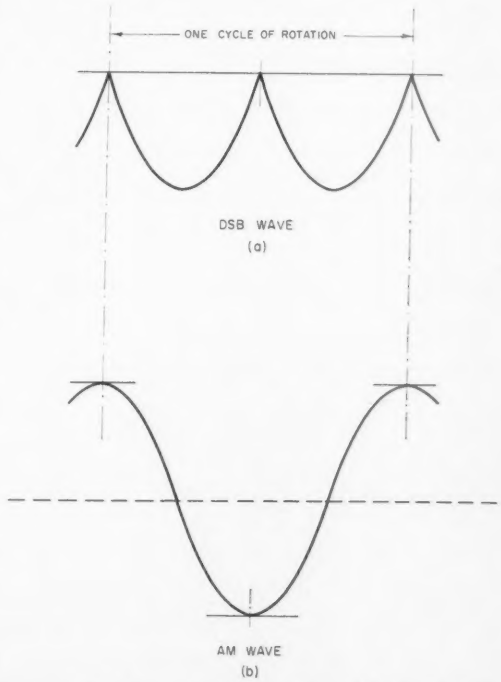


FIGURE 1. *Waveforms for detected DSB and AM mode signals.*

[1] Contribution from Cooley Electronics Laboratory, Electrical Engineering Dept., University of Michigan, Ann Arbor, Mich. This work was supported by the U. S. Army Signal Research and Development Laboratory under Contract No. DA 36-039 sc 78283.

[2] The presentation of this paper was sponsored by the Office of Naval Research. Reproduction in part or in whole is permitted for any use of the United States Government.

[3] Paper presented at the Conference on Transmission Problems Related to High-Frequency Direction Finding, at UCLA, June 21–24, 1960.

In this type of system the bearings are determined by the position of the null of the antenna pattern with respect to a point of reference, usually North. It is seen that in figure 1a there are two such points, thus the familiar 180° ambiguity, while in figure 1b there is only one point at which the signal goes through the null in a specified direction of slope of the wave. It should be noted that the null point has been shifted by decoupling to the dotted line. This brings about an effect, which at present is the most significant detriment to the bearing accuracy obtained, of a shift in the apparent bearing when the character of the receiver is altered even slightly. For example when the gain control is changed over its range the bearing will change by two or three degrees. Since, as explained in a subsequent section, the reconstruction of the visual display and the zero crossing comparison for measuring relative phase both depend on having a constant *average* output from the receiver for a given input signal form but at different levels, it is evident that a receiver is needed that has low envelope distortion in order to eliminate this effect. There is indeed a great need for receivers designed for instrumentation purposes as well as communications, particularly with regard to the phase character of the receiver. If the signal is pulsed then the problem of maintaining the average is more difficult because d–c components are introduced by the pulsing action of the signal. These will average out over a period of time, but for short periods cause havoc in the bearing determination. In viewing the DSB and AM situations certain things come to mind with regard to performing (1) Instrumental measurement of the bearing angle, and (2) Improvement of the signal to noise ratio by applying narrow band synchronous post detection filtering.

In the DSB case the null point region is that of the lowest signal to noise ratio in the entire cycle of rotation; thus in a noisy situation, which is the usual case rather than the exception, the null position is imperfectly defined, and indeed the whole structure of the wave, that is the collection of harmonies which define the structure, is needed in order to determine just where the average null point is located. A similar effect may be had by filtering out the second harmonic term in (4) and noting the relative phase of this wave to a reference wave.

In fact an increase in signal to noise ratio comparable to the AM case may be achieved but since an additional ambiguity is introduced this is not as attractive as using automatic sense if one wants to use narrow band filtering. If it is desired to keep the DSB information as is, except to filter the noise from it, then a comb filter may be used. This may be either the discrete frequency type or more likely the commutated type. In either case it is considerably more involved than a simple narrow-band single frequency filter. In the AM case the phase measurement is made on a sine wave (for a continuous signal) which has the same fundamental period as the goniometer rotation therefore the ambiguity is removed, and one may now apply filtering at the basic scan frequency rather than at twice the scan frequency. A description of the means of obtaining automatic sense, the design of the bearing readout system and a discussion of narrow band post detection filtering with some operational illustrations follow. An important aspect of the design is minimum cost with reasonable performance and simplicity.

2. Automatic Sense

The addition of a carrier or sense signal to the goniometer output changes the receiver input to a standard amplitude-modulated signal, modulated at the spin rate. The phase of this modulation relative to the phase of the rotation gives an unambiguous bearing. It can be shown that if the phase of the injected carrier is within plus or minus 90° of the correct value it distorts the envelope, but only to the extent of adding harmonics (which may be filtered out) and causing a decrease in fundamental amplitude. It does not affect the relative time phase of the fundamental. Thus the generation of the sense signal is not so critical as is often supposed. If the sense signal is larger than the directional signal and is held to within $\pm 45°$ of the proper value, adequate performance may be obtained.

The simplest means for obtaining the sense signal is to insert an antenna in the center of the DF array. Since the goniometer output has suffered some loss, as well as being the difference between two signals, the injected sense voltage can easily be made sufficiently large. If the distance across the array is small compared to a wavelength, a fairly simple network will maintain the phase margin within the required limits [Keen, 1947]. In fact, with rather heavy filtering, satisfactory operation has been obtained in the case illustrated with no additional phasing over the band of operation. The R–390 receiver is conveniently designed with both a balanced and unbalanced input so that the sense signal may be fed directly into the unbalanced antenna input jack.

The instrumentation for producing the proper pattern on the screen of the CRT is depicted in figure 2. This unit serves two functions: (1) it

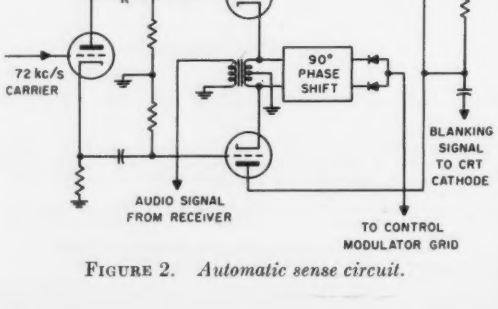


FIGURE 2. *Automatic sense circuit.*

full-wave-rectifies the sine wave output of the receiver to give the usual propeller-type display, (2) it causes a blanking signal to be applied to the grid of the CRT in such a manner that only one half of the propeller appears on the screen. In the AN/TRD–4 equipment the display is filled in with a 72 kc/s carrier. Figure 3 shows the usual input to the indicator goniometer during one complete revolution during DSB type operation. It can be seen that the propeller is described on the face of the tube twice during each revolution. During period T_1 the tube must be blanked during the negative excursions of the 72 kc/s carrier. During period T_2, since the goniometer has rotated 180°, the blanking must occur on the positive excursions if the same half of the propeller is to be displayed. It can be seen from the circuit that with no signal from the receiver both triodes will conduct equally, and, since their inputs are 180° out of phase, cancellation will occur in the plate circuit. In the presence of a signal from the receiver (ideally a sine wave), the tubes will alternately conduct fully and be cut off. It is this alternating operation which causes the proper half of the display to always appear as discussed above. The resultant 72 kc/s signal, when applied to the CRT grid, blanks during the negative excursions and brightens or enhances the desired half of the propeller during the positive excursions.

3. Bearing Readout Computer

The basis for the computer design is that of zero-crossing comparison, as has been used by others [Florman and Tait, 1949; Hatch and Byatt, 1958]. Once the zero-crossing of the reference and DF signals have been determined they may be used to activate either an analog-type coincidence detector or a digital-type bearing computer. It was concluded that for visual monitoring purposes the analog was more suitable, and for detailed processing of the information the direct digital readout was more suitable. For the purpose at hand there was no need to eliminate the 180° ambiguity from the analog readout.

The analog coincidence detector is fed into one channel of a two-channel Brush Recorder. The second channel is used to record signal strength. The particular Brush Recorder used has a frequency response up to about 50 c/s. It is desirable to have an adjustable readout rate in the digital system so that the bearing readout rate can be adjusted to fit the study being made.

For studying the fast-fade variations, Bain [1955] indicates that in order to study the phenomena one should be able to take at least several bearings per second. Of course the rate for slow fluctuation studies should be much less than this, thus there exists the need for variable readout rate.

Reference is made to figure 4, a block diagram of the readout system. The goniometer bearing scale was removed and machined out under the thumb rim, and 1° slits were cut on the outer surface of the rim with an indexing head. Thus, with a light in the

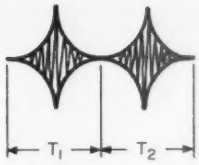


FIGURE 3. *Input to display goniometer during conventional DSB operation.*

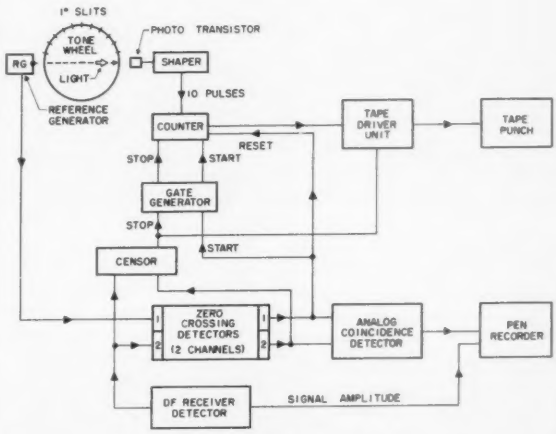


FIGURE 4. *Block diagram of automatic bearing computer and readout.*

wheel and a phototransistor on the outside, a tone wheel is available. A properly adjusted slit in front of the transistor is a necessity. The output of the phototransistor is fed into a two-stage transistor preamplifier and a pulse-forming circuit, then to the counter.

The reference generator is a small permanent-magnet-type alternator which was built on the end of the goniometer and is coupled directly to the shaft. The output from the reference generator is squared and formed into the start pulse in the zero-crossing detector. The output from the DF receiver is similarly treated in the zero-crossing detector unit. The squaring circuitry outputs from these units are fed to the analog coincidence detector and thence to the pen recorder for the visual record of bearing. The pulse output from the reference channel starts the counter, and the pulse output from the receiver stops the counter. The censor unit at present is to be used as a yes-no stop pulse control depending on signal level. The start pulse is delayed slightly in the gate generator in order that the same signal may be used as a reset control at North position. When the censor unit says "no" then the counter will cycle on a 360° count until the censor unit says "yes". The reference point, North in this case, may easily be moved to some other point around the compass if desired. The problem here is that of averaging through the cut point. For example, if the bearings

were symmetrically distributed about North, than the average of these would be South, obviously the wrong answer. One simple solution to this is to have two cut points with two counts and pick the output that is farthest removed from a cut point. The counter output is then fed to the tape punch driver, which is controlled by the censored stop pulse, and thence punched onto paper tape. Since this unit is to be transistorized the Tally 8 channel punch was chosen. Eight channels allow a sector coverage of 255°. In order to cover the full 360° sequential channel coding must be used which still allows a maximum of 30 readings per second.

4. Narrow Band Synchronous Post Detection Filtering

The application of NBSPD filtering has been discussed in the literature [Busignies and Dishal, 1949], however little seems to have been done to exploit its potential to improve bearing sensitivity. In fact, as Busignies and Dishal state, it appears that the capability of utilizing this is one of the fundamental advantages of this type of DF system over the co-operative navigational type system. This stems from the simple fact that in the great majority of situations the determination of the average direction of arrival over period even as short as just a few seconds is a very narrow band process. From an informational point of view the channel capacity required is in terms of bandwidth the approximate frequency of variation of the bearing one wishes to observe. This is of course determined exactly by the reading time of the specific receiving system in use. This means that for most purposes a bandwidth of a few cycles (say less then 10) will be adequate, unless there is some wideband information in addition to the bearing information, or if one is interested in studying the instantaneous rapid fluttering effects caused by wave interference, or for other specialized types of signals. There is more to be gained from a given factor of bandwidth reduction in the predetection circuitry than at post detection. By the same token this is much more difficult to achieve. In applying post detection narrowing if BW (IF) is the bandwidth of the IF in the receiver and BW(A) is the modified audio bandwidth then the following relation gives the input carrier ratio for equal output in terms of bearing determination. This is done for purposes of comparison to the operational illustration shown later.

$$\text{Carrier Ratio}=K\left[\frac{BW(IF)}{BW(A)}\right]^{1/4} \qquad (5)$$

and substituting for BW(IF)=2.0 kc

$$BW(A) = 5.0 \text{ c/s}$$

$$CR=K\left[\frac{2000}{5}\right]^{1/4}=K(400)^{1/4}\cong K(4.475)$$

This means a difference in terms of db of:

$$\text{CR Diff (db)}=20 \log 10\ (4.475)$$

$$=13.0 \text{ db}$$

Results from operation in this manner with automatic sense are given for purposes of illustration in figures 5 through 11. A single tuned circuit with center frequency at the rotational frequency of the goniometer is placed after the detector. This unit has adjustable bandwidth with the narrowest position at about 5.0 c/s. This is the condition under which the pictures were made. This gives a reading time of about 140 msec. which allows observation of bearing changes up to about 7.0 c/s.

Figure 5 shows a strong stable signal in the broadcast band, station WJR in Detroit. The double tip is due to imperfect foldover in the automatic sense circuit, which is being corrected.

In figure 6, (a) and (b) are, respectively, the conventional TRD–4 display and the modified display, in which the target signal, at 6.93 Mc/s, was adjusted in level until one could just estimate a bearing, at around 105°, on the conventional TRD–4 display. These shots are all 4-sec time exposures.

In figure 7, the signal level is reduced by 3.6 db, at which point the TRD–4 display is just barely discernible as a bearing indication.

In figure 8, the signal level is reduced from where it was in figure 6, by 9.6 db, which leaves nothing in figure 8(a), although it can be seen that there is still quite a good bearing indication in figure 8(b). The limit is reached by reducing the signal level by another 6.0 db, producing figure 9. One can still get a fair estimate of the bearing, and the pattern which resembles a one-sided halo, is developing into

FIGURE 5. *Bearing indication on modified AN/TRD–4 for strong stable signal.*

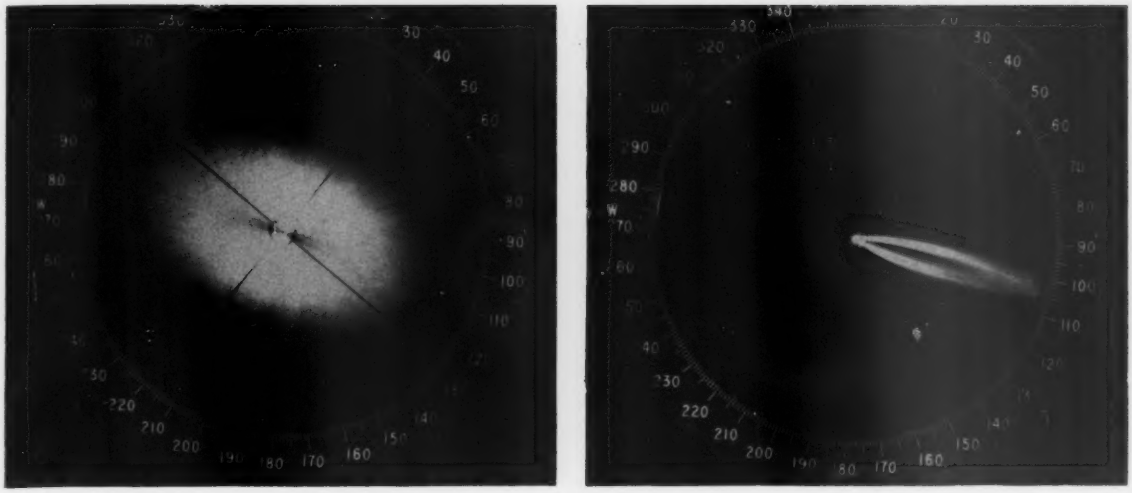


FIGURE 6. *Comparison of operational results from (a) conventional AN/TRD–4 and (b) modified AN/TRD–4.*

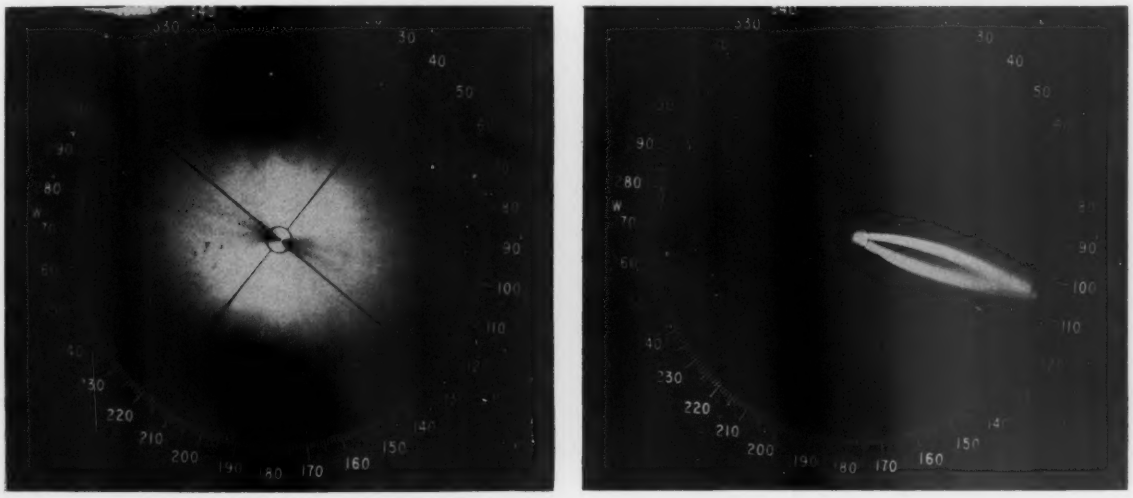
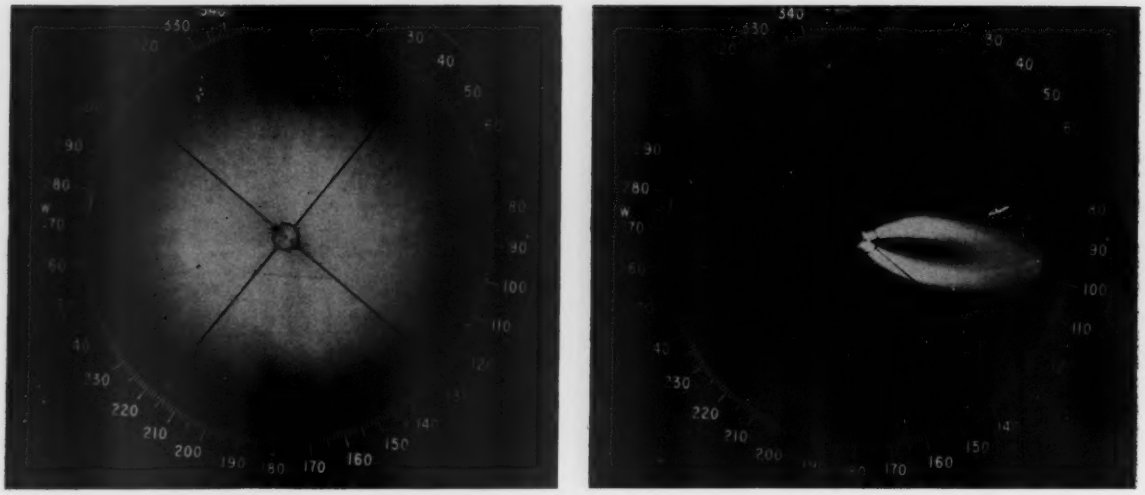


FIGURE 7. *Comparison of operational results from (a) conventional AN/TRD–4 and (b) modified AN/TRD–4.*
Signal Level Reduced by 3.6 db from that in figure 6.



a b

FIGURE 8. *Comparison of operational results from (a) conventional AN/TRD–4 and (b) modified AN/TRD–4.*
Signal Level Reduced by 9.6 db from that in figure 6.

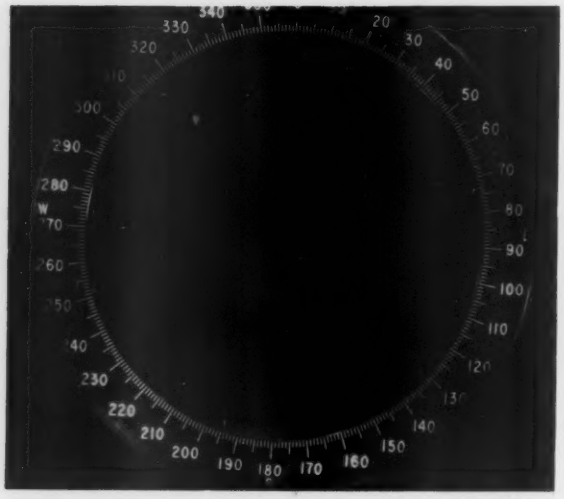


FIGURE 9. *Threshold condition for the modified AN/TRD–4.*

the characteristic background for this mode of operation. With no signal there is a faint green background level on the CRT face. When a signal is tuned in the propeller half suddenly rises out of the hazy background as though by intensity modulation. The difference between the input signal level for figure 6, and that of figure 9, is 15.6 db.

Figure 10, depicts the appearance of a typical, fairly-weak FSK signal at 6.901 Mc/s.

Figure 11, shows the target signal at 6.93 Mc/s. being pulsed at 20 c/s. The integration effect of the film favors 10a more so than the eye in practice.

5. Summary

It has been demonstrated that one may partially automate and otherwise improve the operation of a SGDF with respect to certain prescribed criteria with rather simple and inexpensive methods. One must of course keep in mind what the particular objectives for the system are and the resulting advantages to be gained by making certain changes and the price that sometimes must be paid to make those changes.

6. References

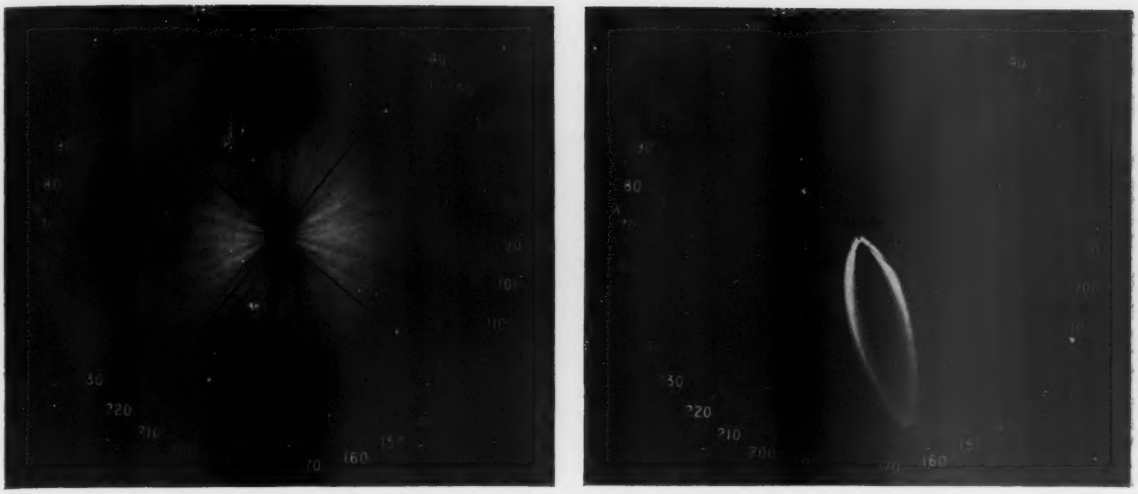
Bain, W. C., On the rapidity of fluctuations in continuous-wave radio bearings at high frequencies, Proc. Inst. Elec. Engrs., Pt. B, **102,** 541–543 (July 1955).

Busignies, H., and M. Dishal, Some relations between speed of indication, bandwidth, and signal-to-random-noise ratio in radio navigation and direction finding, Proc. I.R.E. **37,** No. 5, 478–488 (May 1949).

Florman, E. F., and A. Tait, An electronic phasemeter, Proc. I.R.E. **37,** No. 2, 207–210 (February 1949).

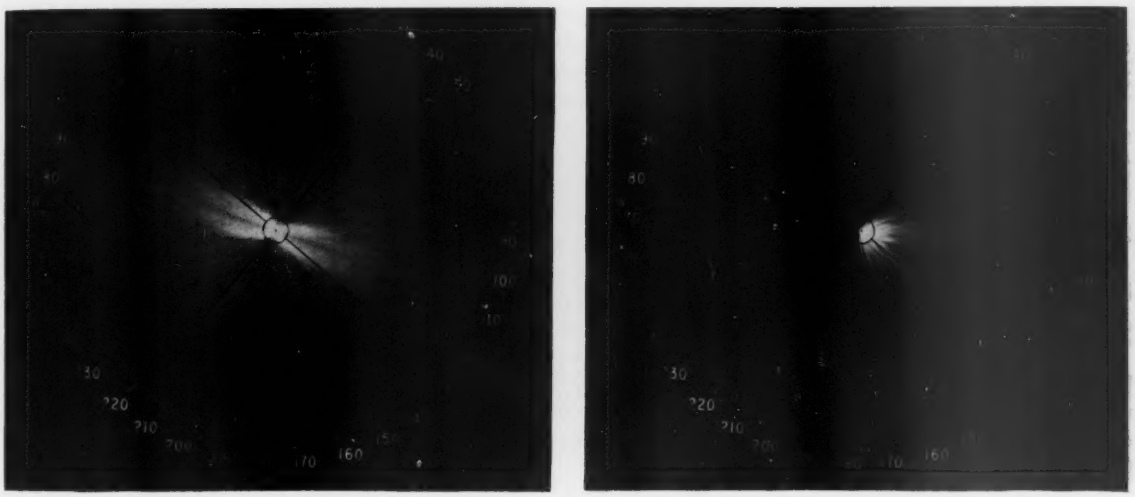
Hatch, J. F., and D. W. G. Byatt, Improvements in HF direction finding by automatic time averaging, Marconi Review **XXI,** No. 128, 16–29 (1958).

Keen, R., Wireless direction finding, pp. 144–149 (Ilife and Sons, Ltd., London, 1947).



a b

FIGURE 10. *Typical, weak FSK signal, as obtained from (a) conventional AN/TRD–4 and (b) modified AN/TRD–4.*



a b

FIGURE 11. *Pulsed signal, as obtained from (a) conventional AN/TRD–4 and (b) modified AN/TRD–4.*

(Paper 65D3–125)

JOURNAL OF RESEARCH of the National Bureau of Standards—D. Radio Propagation
Vol. 65D, No. 3, May–June 1961

Resolution Characteristics of Correlation Arrays [1, 2, 3]

Isham W. Linder

(September 29, 1960)

Antenna arrays which are designed to utilize correlation techniques can result in directivity patterns with very narrow beamwidth. However, analysis of resolution capabilities of these arrays indicates a marked change in expected performance in the presence of two or more signal sources. These effects are analyzed for single frequency signal sources and for randomly varying signal sources. It is shown that optimum results occur when the nonlinear processing of the antenna voltages is limited to a single multiplication. Under these conditions the correlation array has a directivity equivalent to that of a linear array of twice the length.

1. Introduction

Increasing attention has been given in recent years to the merits of correlation techniques as a means of improving the resolution capabilities of passive antenna systems. The possibility of making a decided saving in the overall size of directional antennas, even though this saving must be purchased at the cost of increased complexity of the antenna circuitry, has a distinct appeal in certain antenna applications. It is the purpose of this paper to outline the mathematical analysis of the resolution characteristics of such an antenna array and to compare these results with those which could be expected from a familiar linear additive array.

A correlation array will be defined as one in which voltages induced on the elements are multiplied together and the resulting voltage is averaged over some prescribed time interval to give a desired output voltage. There are, of course, two fundamental variables in this definition. In the first place, when there are a number of elements in the array there are a great many possible combinations of the element voltages. And, secondly, when there are undesired fluctuations in the output voltage, the time interval which is available for averaging this voltage can have a primary influence on the resolution capability of the array.

For this discussion of the mathematical analysis of resolution characteristics the first variable will be avoided by limiting the calculations to specific examples of correlation arrays. The second variable will enter the analysis and will appear in the final results.

It will be assumed that the problem at hand is the resolution of two signal sources separated by some angular displacement θ. Both single-frequency and band-limited signal sources will be considered. The signal sources are assumed to have identical power spectral density over the frequency band of interest. They cannot be resolved by frequency selectivity in the receiving system, but must be resolved by appropriate operations on the voltages induced in the elements of the array.

2. Single-Frequency Sources

In order to analyze the correlation array without becoming too deeply involved in the required calculations, a relatively simple array of four isotropic elements will be first considered. The pattern of the response of this array as a function of θ can be shown to be equivalent to that of a much larger linear additive array. This particular aspect of the capabilities of correlation arrays has been reported in the literature [Berman and Clay, 1957; Jacobson, 1958; Drane, 1959]. These articles have demonstrated that, in response to a single signal source, a correlation array with properly selected element spacing has a pattern which is exactly equivalent to that of a much larger linear array. When two signal sources are present, however, an interaction of voltages produced by the sources occurs in the multiplication processes of the correlator, and this equivalence between linear and correlation arrays has to be reanalyzed.

[1] Contribution from Department of Electrical Engineering, University of California, Berkeley 4, Calif.
[2] This research was supported in part by the Office of Naval Research under Contract Nonr-222(74).
[3] Paper presented at the Conference on Transmission Problems Related to High-Frequency Direction Finding, at UCLA, June 21–24, 1960.

The signal source (see fig. 1) located in the plane of the array and at an angular displacement θ from the main lobe axis, induces a voltage V_i on the i-th antenna element, where

$$V_i = A \cos(\omega t + \tau_i)$$

$$\tau_i = \frac{d_i}{c} \sin\theta \qquad (d_i \text{ is the distance from a reference point in the line of the array to the } i\text{th element.})$$

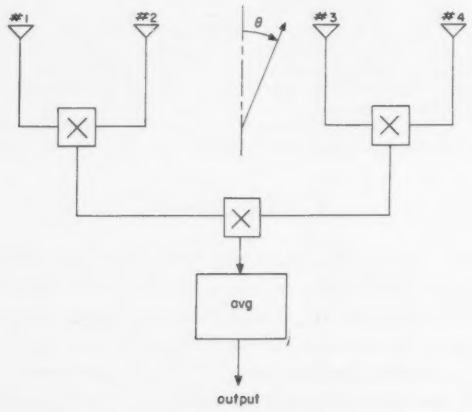


FIGURE 1. *A four-element correlation array.*

If the bandwidth of the integrating circuit is established to reject the second- and fourth-order frequencies resulting from the multiplications, the output voltage is

$$V_{out} = \frac{1}{8} A^4 [\cos\omega(\tau_1 - \tau_2 - \tau_3 + \tau_4) + \cos\omega(\tau_1 - \tau_2 + \tau_3 - \tau_4) + \cos\omega(\tau_1 + \tau_2 - \tau_3 - \tau_4)].$$

And if the relative spacing between elements is established with $d_{12} = D$, $d_{23} = D$, and $d_{34} = 2D$ (and letting $X = \frac{\omega D}{c} \sin\theta$):

$$V_{out} = \frac{1}{8} A^4 [\cos X + \cos 3X + \cos 5X].$$

But this, except for the constant term, is the voltage pattern for a six-element linear additive array with a uniform element spacing of $2D$.

So in this sense the pattern of the four-element correlation array of overall length $4D$ and the voltage pattern of the six-element linear array of overall length $10D$ are equivalent.

Berman and Clay have discussed this equivalence between correlation and linear arrays. In general, directivity patterns for a single monochromatic source have the following mathematical equivalence:

Correlation array	*Linear array*
4 elements, length $4D$	6 elements, length $10D$
6 elements, length $8D$	15 elements, length $28D$
8 elements, length $16D$	32 elements, length $62D$

However, in resolving two similar sources these patterns cannot be used directly, but must be treated with some care. For example, assume source $A(V_A = E \cos\omega_A t)$ is located on the main lobe axis and source B with the same amplitude ($V_B = E \cos\omega_B t$), is displaced through θ^0. The output voltage of the six-element linear array would be:

$$V_{out} = 6E \cos\omega_A t + 2E(\cos X + \cos 3X + \cos 5X) \cos\omega_B t.$$

Whether or not the sources are coherent, the voltage output for arbitrary movement of the sources is simply the sum of the voltage patterns of each source considered individually.

The correlation array, on the other hand, will contain cross product terms resulting from the multiplication processes whose form will depend on the coherence of the two sources and, in general, on the number of multiplication processes.

For the example of the four-element correlation array considered here, the output voltages for two conditions of coherence are·

1. For two coherent sources:

$$V_{out}=\frac{3}{8}E^4+\frac{1}{8}E^4(\cos X+\cos 3X+\cos 5X)+\frac{1}{8}E^4(3+11\cos X$$
$$+10\cos 2X+7\cos 3X+6\cos 4X+3\cos 5X+2\cos 6X).$$

2. For two sources of slightly different frequencies:

$$V_{out}=\frac{3}{8}E^4+\frac{1}{8}E^4(\cos X+\cos 3X+\cos 5X)$$
$$+\frac{1}{8}E^4(4\cos X+4\cos 2X+2\cos 3X+2\cos 4X)+\frac{1}{8}E^4(f(\cos nX\cdot\cos \delta t)$$

where δt is the frequency difference between source A and source B, and $f(\cos nX\cdot\cos \delta t)$ represents a number of terms all containing this beat frequency component.

In both of these equations the first two terms correspond to the linear array pattern, while the remaining terms arise from the nonlinearity of the correlation array. In the case of two sources with identical frequencies, the cross-product terms are constant with time and time averaging of the output voltage will not alter the result. When the sources have different frequencies, however, time averaging can be employed to reduce the beat frequency ($\cos \delta t$) part of the cross-product terms; but even in this case the resulting pattern will not be precisely equivalent to that of the six-element linear array since some of the cross-product terms will remain constant with time.

Calculation of the resolution characteristics of the correlation array is then more involved than the similar calculation for the linear additive array.

Assuming a basic spacing $D=\frac{\lambda}{2}$ the four-element correlation array of length $4D$ will resolve two sources of slightly different frequencies at approximately 19.5°. This is equivalent to the resolution of a uniform linear array with an aperture of $9D$, about twice as long.

A slightly more complex example of a correlation array is one discussed by Drane: The element on the left in figure 2 represents a uniform linear array with aperture a_1, while that on the right is a simple interferometer with aperture a_2.

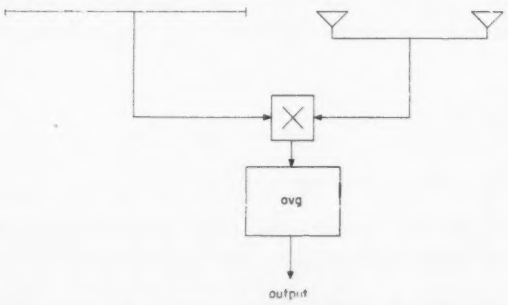


FIGURE 2. *A correlation array employing a linear array and an interferometer.*

The directivity pattern of the uniform linear array is proportional to $\frac{\sin X}{X}$ where $X=\frac{\omega}{c}\left(\frac{a_1}{2}\right)\sin\theta$. If $a_1=a_2$ and there is no spacing between the right end of the linear array and the left element of the interferometer, the directivity pattern of the correlation array is

$$V_{\text{out}} \propto \frac{\sin 4X}{4X}$$

which is the same as that of a uniform linear array of twice the length of the actual array.

However, if two coherent sources are present the output is

$$V_{\text{out}} \propto \frac{\sin 4X_A}{4X_A}+\frac{\sin 4X_B}{4X_B}+\cos\left(X_A+X_B\right)\left[\frac{\sin X_A \cos X_B}{4X_A}+\frac{\cos X_A \sin X_B}{4X_B}\right].$$

The third term again occurs because of the nonlinearity of the correlator.

Calculation of the resolution capability of this array shows that it is equivalent to a uniform linear array one and a half times as long.

These have been only two examples of the effect of cross-product terms in a correlation array. In each case considered, the effect of these terms would vary, depending on the types of signal voltages emitted by the sources and by the number of successive multiplicative processes in the correlator. If the sources were emitting complex signals which could be decomposed into a number of fixed frequency components (a pulse system with a fixed repetition rate, for example) then the cross-product terms resulting from the multiplication would become quite involved.

It should be noted that if the correlator contains only one stage of multiplication, then the cross-product terms occurring from two sources of different frequencies appear only as beat frequency components. These terms occur as a low-frequency a-c signal appearing with the desired d-c measuring voltage. In this case the cross-product terms can be minimized by time averaging the output voltage. However, if more than one stage of multiplication occurs between the antenna element and the output of the array (as was the case in the example of the four element correlation array) the cross-product terms will occur as low-frequency a-c terms and also as d-c terms. In this case the equivalence between the correlation arrays and the linear arrays is not immediately apparent but must be determined by calculation of resolution characteristics for the particular type of signal encountered.

3. Randomly Varying Sources

In those situations in which there is at best only incomplete information about the manner in which the voltage produced by a signal source will vary with time, it is necessary to consider a suitable statistical model which will impose bounds on the expected results and will provide an average description of the voltage variation. The most general such model suitable to this antenna problem is that of a band-limited voltage with a normal distribution. The source can then be described as one which produces a randomly varying voltage whose Fourier series representation becomes:

$$V(t)=\sum_{n=1}^{N} c_n \cos\left(\omega_n t+\Phi_n\right).$$

$V(t)$ is distributed normally with mean zero.

c_n has a Rayleigh distribution with $\overline{c_n^2}=2W(f_n)\delta f$, where $W(f_n)$ is the power density over a frequency interval δf centered at f_n.

Φ_n has a uniform distribution over $(0, 2\pi)$.

ω_o is the lower edge of the band of width $2\Delta f$ c/s and $\omega_i=\omega_o+\frac{2\pi i}{T}$.

The directivity pattern of the correlation array can be stated in terms of the correlation coefficient of the voltage produced by the signal source. For a four-element correlation array the expected value of the output voltage is:

$$\overline{V_1(t)V_2(t)V_3(t)V_4(t)}$$

and, since these are normally distributed voltages, this becomes:

$$\overline{[V_1(t)V_2(t)]}\,\overline{[V_3(t)V_4(t)]}+\overline{[V_1(t)V_3(t)]}\,\overline{[V_2(t)V_4(t)]}+\overline{[V_1(t)V_4(t)]}\,\overline{[V_2(t)V_3(t)]}$$

$$=R(\tau_{12})R(\tau_{34})+R(\tau_{13})R(\tau_{24})+R(\tau_{14})R(\tau_{23})$$

where $R(\tau_{ij})$ is the correlation coefficient of the voltages induced on elements i and j.

If the signal source is band limited, $f_c\pm\Delta f$, and the signal has a uniform power spectral density over this pass band, then

$$R(\tau)=R(0)\,\frac{\sin 2\pi\Delta f\tau}{2\pi\Delta f\tau}\cos 2\pi f_c\tau. \qquad \text{(autocovariance)}$$

With proper spacing of the correlation array elements this again can be put in a form resembling the directivity pattern of a uniform linear array.

$$\overline{V_1(t)V_2(t)V_3(t)V_4(t)}=3R(O)^2[A\cos X+B\cos 3X+C\cos 5X]$$

where $X=2\pi f_c\,\frac{D}{c}\sin\theta$ and the coefficients A, B, and C are combinations of $\frac{\sin 2\pi\Delta f\tau}{2\pi\Delta f\tau}$ terms which approach the value 1 as the receiver pass band is decreased.

If there are two such sources separated by an angular displacement θ, the output voltage of a linear array could be suitably employed to resolve the sources. An instantaneous voltage observation at various positions as the array is rotated past the sources would generally not be an adequate procedure due to the assumed randomness of the sources. Some sort of time averaging procedure (as in an average power measurement) or a procedure requiring a definite observation interval (the peak voltage occurring in the interval, for example) would be necessary.

As an example of this use of a linear array, assume that the output of the array will be squared and resolution obtained by power measurement. As before, assume there are two sources, source A on the main lobe axis and source B displaced through $\theta°$. Then before squaring:

$$V_{\text{out}}=6V_A+2V_B(\cos X+\cos 3X+\cos 5X)$$

$$V_{\text{out}}=6\sum_{n=1}^{N}c_n\cos(\omega_n t+\phi_n)+2\sum_{n=1}^{N}d_n\cos(\omega_n t+\phi_n)(\cos X+\cos 3X+\cos 5X)$$

After squaring:

$$V_{\text{out}}^2=36V_A^2+24V_AV_B(\cos X+\cos 3X+\cos 5X)+4V_B^2(\cos X+\cos 3X+\cos 5X)^2$$

$$V_{\text{out}}^2=36\left[\sum_{n=1}^{N}\sum_{m=1}^{N}c_nc_m\cos(\omega_n t+\phi_n)\cos(\omega_m t+\phi_m)\right]$$

$$+24\left[\sum_n\sum_m c_nd_m\cos(\omega_n t+\phi_n)\cos(\omega_m t+\phi_m)(\cos X+\cos 3X+\cos 5X)\right]$$

$$+4\left[\sum_n\sum_m d_nd_m\cos(\omega_n t+\phi_n)\cos(\omega_m t+\phi_m)(\cos X+\cos 3X+\cos 5X)^2\right]$$

The first and last terms of this expression have χ^2 distributions with mean values equal to the power from the signals; the middle term has a bivariate normal distribution with mean zero. So the average value of the squared voltage term will be equal to the power received from the two sources. The instantaneous value will fluctuate about the average, and time averaging can be used to minimize the effect of the fluctuation.

An expression for the effect of an averaging device can be obtained from general filter considerations [Middleton, 1960]. If $h(t)$ is the effective weighing function of a linear measuring device and $x(t)$ is the function to be measured, a measurement $M_x(T)$ made at time $t=T$ after $x(t)$ has been introduced at $t=0$ can be expressed by the convolution

$$M_x(T)=\int_0^T h(u)x(T-u)du \qquad h(u)=0,\ u<0$$

where $(0,T)$ is the observation interval. $M_x(T)$ will vary from observation to observation, fluctuating about the expected value $\overline{M_x(T)}$ with a variance $\sigma^2=\overline{M_x(T)^2}-[\overline{M_x(T)}]^2$.

In the general situation, $x(t)$ is at least wide sense stationary, and

$$\overline{M_x(T)}=\int_0^T h(u)\overline{x(T-u)}du$$

$$\overline{M_x(T)^2}=\int_0^T\int_0^T h(u)\overline{x(T-u)x(T-v)}h(v)du\,dv$$

$$=\int_0^T\int_0^T h(u)\psi_x(u-v)h(v)du\,dv$$

where ψ_x is the autocovariance coefficient of the function to be measured. For an ideal integrator $h(u)=\frac{1}{T}$, $0<u<T$. And

$$\overline{M_x(T)}=\overline{x(t)}$$

$$\overline{M_x(T)^2}=\frac{2}{T}\int_0^T\left(1-\frac{s}{T}\right)\psi_x(s)ds$$

The correlation coefficient of the power function can be obtained in terms of the correlation coefficient of the voltage. The voltage has a normal distribution; the correlation coefficient of the power.

This is, of course, a direct indication of the fluctuation of the output power about the mean value. For this case of two statistically similar sources, the variance determines, in terms of confidence intervals, the difference in mean power which the array receives when it is directed at one of the sources and when it is directed between the sources in order for the sources to be consistently resolved.

Now, with this description of the resolution process by a linear array it is possible to investigate the resolution process in a correlation array and to compare the resultant effects. As in the previous case, it is necessary to determine the mean value of the output voltage and to determine the correlation coefficient of this voltage in order to establish bounds on the expected fluctuation of the measurement.

The expected value of the voltage produced at the output of a four-element correlation array by two independent sources (source A and source B) can be calculated either directly or by the characteristic function method and can be expressed in terms of the correlation coefficients of the individual voltages.

With two sources present, the voltage on the ith element becomes:

$$Vi(t)=V_A(t+\tau_i)+V_B(t+\tau_i).$$

and

$$\overline{M_x(T)}=\overline{V_1(t)V_2(t)V_3(t)V_4(t)}=\{[R_A(\tau_{12})R_A(\tau_{34})+R_A(\tau_{13})R_A(\tau_{24})+R_A(\tau_{14})R_A(\tau_{23})]$$
$$+[R_B(\tau_{12})R_B(\tau_{34})+R_B(\tau_{13})R_B(\tau_{24})+R_B(\tau_{14})R_B(\tau_{23})]+[R_A(\tau_{34})R_B(\tau_{12})+R_A(\tau_{12})R_B(\tau_{34})$$
$$+R_A(\tau_{24})R_B(\tau_{13})+R_A(\tau_{23})R_B(\tau_{14})+R_A(\tau_{14})R_B(\tau_{23})+R_A(\tau_{13})R_B(\tau_{24})]\}.$$

$\overline{V(t)^2V(t+s)^2}$, will be the coefficient of $\frac{(iu)^2}{2}\frac{(iv)^2}{2}$ in the expansion of the characteristic function of $V(t),V(t+s)$.

$$C(u,v,s)=\overline{\exp\,[iuV(t)+ivV(t+s)]}=\exp\,[-\tfrac{1}{2}R(0)(u^2+v^2)-R(s)uv]$$

where $R(s)$ is the correlation coefficient of the voltage. Then

$$\psi(s)=R(0)^2+2R(s)^2.$$

If it is assumed that the receiver circuits have a rectangular pass band, $f_c\pm\Delta f$, and that the power spectral density of the signal sources is uniform over this pass band, then

$$R(0)=\text{average power from the signals}$$

$$R(s)=R(0)\,\frac{\sin 2\pi\Delta fs}{2\pi\Delta fs}\cos 2\pi f_o s.$$

Substituting these values into the equations describing the ideal integrator:

$$\overline{M_x(T)}=R(0)$$

$$\overline{M_x(T)^2}=\frac{2}{T}\int_0^T\left(1-\frac{s}{T}\right)(R(0)^2+2R(s)^2)ds$$

$$=[\overline{M_x(T)}]^2\left[1+\frac{4}{T}\int_0^T\left(1-\frac{s}{T}\right)\left(\frac{\sin 2\pi\Delta fs}{2\pi\Delta fs}\right)^2(\cos 2\pi f_c s)^2ds\right]$$

which, for large values of $2\pi\Delta fT$, becomes

$$\overline{M_x(T)^2}=[\overline{M_x(T)}]^2\left(1+\frac{\pi}{2\pi\Delta fT}\right)$$

The variance of this averaged power then is

$$\sigma^2=\frac{\pi}{2\pi\Delta fT}\,[\overline{M_x(T)}]^2=\frac{\pi}{2\pi\Delta fT}R(0)^2.$$

The first and second terms of this expression give the expected voltage due to each source individually. The third term contains the cross product components which occur as a result of the two stages of multiplication that the element voltages undergo.

The variance of this output voltage could be calculated directly; however, the large number of terms in the final expression would make this quite laborious. A good approximation to this variance can be made by calculating the variance of the voltage produced by a single source located on the main lobe axis. The correlation function of this voltage can be determined readily by the characteristic function method:

$$\psi(s)=9\,R(\mathrm{O})^4+72\,R(\mathrm{O})^2R(s)^2+24\,R(s)^4.$$

If we again assume an ideal integrator and a band limited process, the variance becomes:

$$\sigma^2=\frac{14\pi}{3(2\pi\Delta fT)}\,[\overline{M_x(T)}]^2. \qquad (2\pi\Delta fT)>>1.$$

This variance is seen to be somewhat more than four times that found for the single multiplication required in the power calculation. And, in general, if more elements are added to the correlation array, this variance will increase approximately by a factor of four for each pair of elements added to the array.

This increase in variance indicates an increase in the fluctuation component of the output voltage as the number of multiplicative processes are increased. This results in an increase in the integration time required to reduce this fluctuation to some prescribed level, and, therefore, makes this time averaging interval an important part of any discussion of the resolution capabilities of these arrays.

Band-limited random signals

Resolution at the 95% confidence level

Four-element correlation array length 4D		
Resolution	Integration time	Aperture of equivalent uniform linear array
19°	$2\pi\Delta fT=$ 56	9.6D
18°	$2\pi\Delta fT=$ 303	9.4D
17.5°	$2\pi\Delta fT=$2790	9.2D

$$D=\frac{\lambda}{2}$$

4. Summary

An analysis of the expected performance of antenna arrays which utilize correlation techniques indicates the possibility of a marked saving in antenna size. In the general situation, the multiplicative processes in the correlator will introduce cross-product terms which will appear as low-frequency fluctuations in the antenna output voltage. Additionally, these cross-product terms may contribute to the d-c output voltage. The appearance of the cross-product terms complicates the calculation of resolution capabilities, and generally necessitates the definition of a time-averaging interval in any discussion of this resolution. In the examples considered here, it was evident that the directivity pattern alone did not describe the performance of a correlation array in resolving two signals. In these examples the relatively simple correlation arrays had resolution capabilities equivalent to those of uniform linear arrays of about twice the length.

5. References

Berman, A., and C. S. Clay, Theory of time averaged product arrays, J. Acoust. Soc. Am. **29,** No. 7, p. 805 (July 1957).

Drane, Charles, Phase modulated antennas, Air Force Cambridge Research Center, Tech. Rep. 59–138 (April 1959).

Jacobson, M. J., Analysis of a multiple receiver correlation system, J. Acoust. Soc. Am. **29,** No. 12, p. 1342 (December 1957); **30,** No. 11, p. 1030 (November 1958).

Middleton, David, An introduction to statistical communication theory, Chapter 16 (McGraw–Hill, New York, N.Y., 1960).

(Paper 65D3–126)

JOURNAL OF RESEARCH of the National Bureau of Standards—D. Radio Propagation
Vol. 65D, No. 3, May–June 1961

Instrumentation for Propagation and Direction-Finding Measurements[1, 2]

Edgar C. Hayden

(August 26, 1960; revised January 3, 1961)

Limitations imposed on radio direction-finding systems are discussed in terms of a generalized representation of such systems in the form of a block diagram. Factors affecting these limitations include: (1) considerations of signal-to-noise ratio in the early part of the system, (2) receiver bandwidth requirements for adequate selectivity, (3) width of spectrum generated by prereceiver encoding and computing processes, (4) restriction to linear processes in multisignal portions of the system, and (5) availability of operational devices suitable for use in low-signal-level portions of the system.

A generalized representation of a radio direction-finding system can be given in the form of a block diagram or flow chart (fig. 1) showing the operations or functions which are necessary in translating the basic measurements on the incident signal field into the values of the desired incident field parameters. Limitations are imposed on the system at each stage in the process. Factors leading to these limitations include: (1) Considerations of signal-to-noise ratio in the early part of the system, (2) receiver bandwidth requirements for adequate selectivity, (3) width of spectrum generated by prereceiver encoding and computing processes, (4) restriction to linear processes in multisignal portions of the system, and (5) availability of operational devices suitable for use in low-signal-level portions of the system.

The development of the detailed configuration of any direction-finding system is based on some concept of the character of the incident signal field. This character is described in terms of a mathematical model, the parameters of which are the quantities to be evaluated by the system. To this point, the incident signal field models which have been used as bases for direction-finding system development have been rather simple, far too simple in fact to satisfactorily describe the actual signal fields. One of the frontiers in direction-finding system development lies, at the moment, in the area of improving the incident field model and of developing more sophisticated systems based on the improved signal field models.

Another frontier lies in the area of applying modern computing techniques to direction-finding system design. One of the major functions of a direction-finding system is, after all, the performance of a computational operation. To this point all systems in use are based on some simple-minded analog computation scheme. This is an unnecessarily restrictive limitation, especially considering the tremendous strides which have been made in recent years in the field of electronic computation.

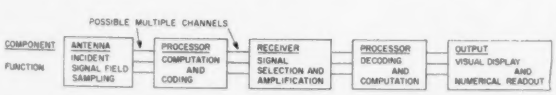


FIGURE 1. *Generalized direction-finding system.*

Two specific direction-finding systems are proposed, both intended primarily for use with antenna arrays of wide aperture. In one system, intended principally for circular arrays, a commutating device is used to encode the output of the antenna array in a form suitable for transmission through a single receiver channel. The output of the receiver would be analyzed, using sampled data techniques, to obtain the required bearing information. The second system makes use of a twin-channel receiving system to avoid the necessity for relatively slow scanning, thus permitting essentially instantaneous acquisition of bearing data (by "instantaneous" is meant a time small enough to permit obtaining a bearing on the shortest pulse transmittable through the receiver). The unusual feature of both systems is the use of an electronic digital computer to perform the bearing interpolation between planes of symmetry of the array, and to provide the final numerical output of the system. The conventional "indicator" is retained in modified form as an operator aid, and to permit the operator to "censor" information fed to the computer, or to "backstop" the computation system in situations where it would fail.

NOTE: A more detailed discussion of the material presented in this paper will be available in printed form during the first quarter of 1961. It will appear in *Interim Engineering Report No. 9, Wullenweber Direction-Finding System* to be published by the Radio Direction-Finding Research Laboratory, Department of Electrical Engineering, University of Illinois, Urbana, Illinois, for the Bureau of Ships, under Contract Nobsr 64723.

(Paper 65D3–127)

---

[1] Contribution from Radio Direction Finding Research Group, Department of Electrical Engineering, University of Illinois, Urbana, Ill.
[2] Summary of paper presented at the Conference on Transmission Problems Related to High-Frequency Direction Finding, at UCLA, June 21–24, 1960.

JOURNAL OF RESEARCH of the National Bureau of Standards—D. Radio Propagation
Vol. 65D, No. 3, May–June 1961

Brooke Variance Classification System for DF Bearings [1, 2]

E. M. L. Beale

(August 22, 1960; revised December 27, 1960)

This paper describes the advantages of having an objective classification system for DF bearings. The Brooke system is described in some detail, and the problems involved in setting up a system on these lines are considered.

1. Introduction

For many years we used a simple ABCD classification system for HF DF bearings taken with U-Adcock direction finders. This was not entirely satisfactory for various reasons. The fundamental difficulty was that, although in theory the classification depended in a well-defined way on the consistency of the observed bearings, in practice it was largely subjective. The Brooke Variance Classification System is based on the Ross-Barfield system developed during the Second World War and described by Ross [1947]. The present paper is an attempt to describe the system from the point of view of a group that might be considering the introduction of such a system.

Section 2 of this paper reviews the reasons for having a bearing classification system. The sources of error in DF bearings are described briefly in section 3. The history of the Brooke system is reviewed in section 4. Section 5 is devoted to a discussion of the problems involved in introducing a similar system into another network. The solutions found for the Brooke system are indicated. Section 6 describes the statistical analysis required to estimate variance components from check bearing data.

2. Why Does One Need a Classification System?

The first question is, why does one need a bearing classification system? It is impossible to give a short answer to this question, since it depends on the answer to an even more fundamental question: What use is one trying to make of one's DF bearings? One will presumably plot the bearings reported on any task on a gnomonic chart. One can then either

(1) gaze in wonder at the set of plotted lines, and finally emerge with a point estimate for the position of the transmitter, and a circle of arbitrary radius centered on this point such that the true position "probably" lies within the circle,

or (2) use one of the many unsound plotting methods that have been proposed from time to time,

or (3) use an electronic computer to find a more precise least squares solution, i.e., to find

(a) a Best Point Estimate (or Most Probable Point) defined as the point minimizing the weighted sum of squares of the angular errors, the weights being inversely proportional to the assumed variances of bearing errors at the station concerned, and

(b) a region, normally a circle or a rectangle, approximating to the region where this weighted sum of squares exceeds its minimum value by less than some given constant, such as 4. The region may be called a 90 percent (say) Probability Region, or Confidence Region. Its purpose is to give an appreciation of the probable accuracy of the point estimate,[3]

or (4) use a method based on some valid approximation to the least squares solution.

In cases (1) and (2) it may be helpful to have some rough measure of the accuracy of any particular bearing, but the traditional ABCD classification based on the consistency of repeated observations, or some other subjective method, may be good enough.

In cases (3) and (4) it is important to have a realistic measure of the probable accuracy of any particular bearing. Ross [1947] wrote

> "It has long been customary for bearings to be classified somewhat arbitrarily by the direction-finding operator. This classification, intended to give an indication of the probable reliability of the bearings, is usually based on the quality (sharpness) and steadiness of the minimum. This practice is a legacy from the days when radio communication was conducted chiefly on the long and medium wavelengths, and it was not unreasonable to expect the operator to estimate the reliability of bearings in this way. In the short-wave band, however (20 to 100 m approximately), where propagation is chiefly controlled by ionospheric reflections, conditions are far from simple. Bearings are generally in a state of continuous variation and it becomes increasingly difficult for an operator to classify them by any more or less intuitive process."

[1] Contribution from Admiralty Research Laboratory, Teddington, Middlesex, England.

[2] Paper presented at the Conference on Transmission Problems Related to High-Frequency Direction Finding, at UCLA, June 21–24, 1960.

[3] The conventional justification for this as a Confidence Region is based on the assumption that all possible bearing lines from a given station are approximately parallel. Beale [1960] provides a method of justifying it even when this approximation is unacceptable. Beale's paper shows that the approximate confidence region is justified if a quantity N_ϕ, called the "intrinsic nonlinearity of the model" is smaller than about 0.1. It seems likely that the DF problem nearly always satisfies this condition; even though one can easily imagine situations with a "narrow base line" when the problem is decidedly nonlinear when expressed in terms of natural parameters, such as the latitude and longitude o the transmitter.

Ross proceeds to describe an objective classification system which is the basis of the Brooke system. The details of such systems will be described in section 4. Meanwhile we should consider another advantage in having a rational and objective classification system: it is almost indispensable to a thorough quantitative analysis of the magnitudes of the various errors arising in practical DF. It is virtually impossible to get a set of observations that is statistically balanced with respect to all relevant factors. So, for example, when studying the effect of distance one must allow for the difference in frequencies of transmitters at different distances; and such a process requires the estimation of variance components due to the different errors making up the observed bearing error.

3. Sources of Error in HF DF Bearings

The sources of error in HF DF bearings are discussed by Ross [1947] and Bowen [1955]. They can be summarized as follows:

(1) Instrumental errors due to small errors of antenna balance, and similar causes. These errors can be greatly reduced by a local calibration, but there may be an appreciable residual error varying with radio frequency and direction of incidence (in azimuth and elevation) and from site to site.

(2) Distant-site errors, due to irregularities of, and obstacles on, the terrain around a high-frequency direction finder but outside the near site or calibration area. The values of these errors vary with radio frequency and direction of incidence (in azimuth and elevation). The variances vary from site to site, and tend to decrease as the radio frequency increases.

From our present point of view, errors of types (1) and (2) are similar. They are virtually independent of time. For a given transmitter working on a given frequency the errors may vary a little in that the angle of elevation of the incoming ray will depend to some extent on ionospheric conditions, but there will be an appreciable systematic component of such an error.

(3) Ionospheric lateral deviation errors, due to large scale irregularities in the ionosphere. These errors vary only slowly with time (with periods of the order of tens of minutes). Their variance does not depend on the DF site, but depends strongly on the distance of the transmitter. Not much is known about the effects of time of day and radiofrequency on lateral deviation. It seems possible that errors are somewhat greater at night than by day and somewhat greater on low frequencies than on high frequencies. If so, this is probably due to the increased number of multiple reflections possible at night and on lower frequencies. Large systematic and random effects can be observed around sunrise. Later studies have not essentially changed the views expressed by Ross [1949] on the subject of lateral deviation.

(4) Wave interference errors and polarization errors. These errors may be quite large, but they generally vary rapidly (with periods of the order of seconds). Their effect can therefore be greatly reduced by taking the average bearing (preferably weighted according to signal strength) over a period of 10 sec or longer.

4. History of the Brooke System

The Brooke Variance Classification System arose out of the system developed during the Second World War by Mr. W. Ross and Dr. R. H. Barfield, and described by Ross [1947]. The basic assumption underlying the system is that the error in any observed DF bearing is a sum of three uncorrelated errors:

(1) An instrumental and distant site error, whose mean square value can be determined as a function of the radiofrequency and the DF site.

(2) A lateral deviation error, whose mean square value can be determined as a function of the distance from the transmitter to the DF site. This component must of course be estimated at the plotting center.

(3) A wave interference and polarization error. An estimate of the mean square value of this error can be determined from the mean "swing" of the observed bearing over a period of 1 or 2 sec, or from the scatter of repeated observations made at intervals of about 5 or more sec, or both. The distance from the transmitter to the DF site can also be used to throw some light on the probable magnitude of this error, since wave interference due to multipath arrival is apt to be more serious on the more distant targets.

When after the war attention was again directed to the classification of HF DF bearings, the Ross system was examined but considered to be unsatisfactory in its existing form. The users also considered the subjective method of classifying bearings, then in use, to be unsatisfactory. Mr. Norman Brooke, while a member of the Admiralty Department of Operational Research, carried out an investigation into the problem of classifying HF DF bearings, the outcome of which was a proposal to modify the Ross system, and to extend it to apply to visual, as well as aural null, equipment. He found that it was necessary to give the variance components values that were rather larger than those proposed by Ross.

The modified version was found to be satisfactory. It was decided, with Mr. Ross's concurrence, that the system should be called the Brooke system.

Before the system was introduced a statistical analysis was made of a large number of check bearings on known transmitters, in order to derive realistic estimates of all the variance components. Further, it is known that station performance changes to some extent with time, so the analysis of check bearing data continues on a routine basis in order to keep track of these changes. It is found that the variance component depending on the DF station has to be changed by 1 unit (i.e., 1° squared) from time to time—perhaps once every 2 years on the average. The other variance components are

reviewed from time to time, but changes have not been found necessary.

So we may conclude that the Brooke system works satisfactorily. The essential reason for this seems to be that the system combines physical theory and empiricism in about the right proportions for our present state of knowledge about HF DF. We certainly do not maintain that no further improvements in the system are possible.

5. How Could a Similar Classification System Be Introduced Into Another Organization?

Suppose that some authority has decided that it would be worth introducing an objective bearing classification system on these lines into his HF DF network, what steps must be taken to implement this decision?

The first task is to define a set of factors on which the variance is assumed to depend. It would seem natural to include

(a) The DF station,
(b) the radiofrequency,
(c) the distance of the transmitter from the station, and
(d) some measure of the quality of the observed bearing.

Some discussion on this last item is perhaps required. A cursory reading of Ross [1947] might give the impression that an objective classification system is only possible if one takes repeated snap observations to determine the bearing. This is not so. The true situation is as follows:

Admittedly if one takes repeated snap observations (which may be mental averages over a period of 1 or 2 sec), then one is likely to obtain a more accurate bearing than if one simply gazes at the display for about 30 sec, because experience has shown that it is almost impossible to average mentally over a period of more than about 5 sec. Furthermore, the scatter of the individual snap bearings provides additional information concerning the reliability of the mean bearing. But if the transmitter is not active for long enough to take repeated observations, or if the bearing is required so urgently that there is not time to take and report the results of several snaps, then one can still use an objective classification system based on other factors. Indeed both the original Ross-Barfield system and the Brooke system include means of assessing a variance when it is impossible to obtain more than one snap.

The next task is to postulate a formula for the variance as a function of these factors with a finite number of unknown parameters.
It seems natural to write

$$B=V_f+V_d+V_q,$$

where B (for Brooke) denotes the total variance (i.e., the total mean square error if one proceeds as if all errors had zero means).

V_f denotes a component depending on radiofrequency and the DF station, representing instrumental and distant site errors,

V_d denotes a component depending on the distance from the transmitter, representing lateral deviation error, and V_q denotes a component depending on the quality of the observed bearing, and the type of equipment and display used, representing wave interference and polarization error, and also observational error due to imperfections in the DF operator.

We must still specify the forms of V_f, V_d, and V_q to reduce them to a finite number of parameters. It may be helpful to consider each in turn with reference to the existing Brooke system.

The Brooke system assumes that

$$V_f=a_j+b_jf,$$

where a_j and b_j are parameters estimated separately for each DF station, and f denotes the frequency band. The definition of these frequency bands is being reconsidered. There is some evidence that the variance is not very sensitive to radiofrequency above about 9 Mc/s. The following bands may therefore prove appropriate.

$f=1$ for frequencies between 2.000 and 3.999 Mc/s,
$f=2$ for frequencies between 4.000 and 8.999 Mc/s,
$f=3$ for frequencies above 9.000 Mc/s.

For the distance component, the Brooke system assumes that

$$\begin{aligned} V_d&=0.6/d^2 \text{ for } d<0.8,\\ V_d&=1 \quad\ \text{for } 0.8\leq d<3.5,\\ V_d&=3 \quad\ \text{for } 3.5\leq d, \end{aligned}$$

where d denotes distance from the transmitter to the DF station, measured in thousands of kilometers.

For the quality component, the Brooke system assumes that

$$V_q=V_s+V_n,$$

where

$V_s=0$ for swings between 0 and 8°,
$=1$ for swings between 9 and 13°,
$=2$ for swings between 14 and 18°,
$=4$ for swings between 19 and 23°,
$=6$ for swings between 24 and 37°,
$=9$ for swings over 38° or unmeasured,
and $V_n=r^2/nd_n^2$,

where n is the number of snap bearings taken, r is the range of these snaps, and d_n is the average range of n observations from a Gaussian population with unit variance (tabulated by Tippett [1925]).

The theory underlying the use of these formulas is that V_n might be expected to give a valid estimate of the variance of those components of error (due to polarization errors and wave interference) that vary rapidly. But in practice the operator cannot help

being biased by the knowledge that all his snaps refer to the same transmission. He will therefore tend to make his observations more consistent than they should be, and an additional variance component must be introduced to allow for this. The bias is likely to depend on the difficulty of the task of taking a bearing, which is measured by the average "swing". For an aural null equipment the swing is defined as the width of the arc about the minimum over which there is a just detectable change of signal. For any equipment with a visual display it is defined as the angle through which the instantaneous value of the bearing moves during the time taken to read its mean value for an individual snap. There is of course no logical reason why the formula for V_s should be the same for different types of display, even though all the equipments have the same antenna system. But the statistical analysis of check bearings suggests that there is no important difference in this respect between aural null, spinning goniometer, and twin-channel c.r.d.f. displays.[4]

If only one snap is taken, increased values of V_d and V_s are used to compensate for the absence of the V_n term.

Different organizations may find it convenient to estimate V_q differently. This must depend on what information the DF operator finds it convenient to report. If he simply reports a letter classification (A, B, C, or D), then one could put $V_q=0$ for an A class bearing, and estimate 3 constants, being the values of V_q for B, C, and D class bearings respectively.

One might also use the signal strength as an indicator of bearing quality. This does not in fact enter into the existing Brooke system.

Having defined the form of the bearing classification system, one must estimate the parameters. One can get preliminary estimates of these from physical theory and from the values used in existing systems, but these estimates must be corrected (or confirmed) using data obtained by the DF organization in question. These data may be either genuine operational data or check bearing data where the true positions of the transmitters are known to the analyst.

As far as I know, no serious attempt has yet been made to estimate variances from operational data when the true positions of the transmitters are unknown. The use of such data poses the fundamental difficulty that one has to lean rather heavily on the assumption that the bearing errors are uncorrelated. If one is prepared to accept this assumption, then it is fairly easy to estimate the absolute values of the variances of a set of DF stations if we know their relative variances. For in this case we can determine the true Best Point Estimate; i.e., the hypothetical position of the transmitter minimizing the sum of squares of the bearing errors divided by their variances. We can then use the fact that (if the error distributions are normal with zero mean) the minimum value of this sum is distributed as χ^2 with n-2 degrees of freedom, where n is the number of DF stations contributing to a task. In particular, the mean value of this sum is n-2.[5]

In practice the relative variances are unlikely to be known. The problem then becomes considerably more complicated. It is discussed in a companion paper, Beale [1961].

It is much easier to estimate variances from check bearings, provided satisfactory check bearings can be obtained.

Check bearings may be used both to verify that the equipment is functioning properly and to indicate the probable accuracy of the DF station on operational tasks. A check bearing program to check the equipment does not need to be controlled nearly as carefully as a check bearing program to estimate accuracy. Bowen [1955] writes

> "The use of check bearings on known transmitters in establishing the performance of a high-frequency direction-finder is a widely accepted technique. It has been found, however, that two problems present themselves:
>
> (a) There is difficulty in arranging a programme in which the radio frequencies, ranges, azimuths, propagation paths, transmitter powers and other parameters are adequately sampled and distributed similarly to those which will define the normal task of the direction finder.
>
> (b) Unless great care is taken to hide the identity of the transmission from the operator, both as regards its true bearing and its identity with previous transmissions, considerable operator bias will occur.
>
> "The methods by which these problems are overcome will generally be peculiar to the task with which the d.f. organization is concerned. It may be that (a) will require the deliberate setting-up of a transmitting system, although careful selection of known transmitters may obviate this; (b) can be overcome either by careful organization of the way in which check-bearing tasks are fed to the operator, or by a system of random scale displacements at the direction-finder."

Even if one fails to overcome these problems completely, it would seem unreasonable to use this as an excuse for not introducing an objective classification system. Indeed the results from an objective system should be very much better than those derived from overall station performances based on such data—since the system will give some rational method of extrapolating to conditions met only rarely in the check bearing program.

The statistical problems involved in estimating variance components from check bearings are discussed in the next section.

[4] There is a small point of some statistical interest about the estimation of V_q. Elementary theory suggests that $\sqrt{V_n}$ is an unbiased estimator of the standard deviation of errors varving randomly from snap to snap, for a given value of this standard deviation. But this does not prove that $\sqrt{V_n}$ provides the best estimate of this standard deviation, given the range r. Indeed one might expect that a better estimate would be obtained by attaching some weight to the estimate obtained from the long-term average value of this standard deviation—thereby making the used estimate less dependent on r. After holding forth on these lines to my colleagues on more than one occasion, I eventually persuaded them to do a detailed analysis of the check bearing data to investigate this point. They found that, if anything, the variance component V_n should be made more dependent on r. It has been suggested that this is because the dividing factor d_n^2 is too large because the error distribution is not normal. (And indeed even from a normal population a smaller dividing factor is required to give an unbiased estimate of the *variance* as opposed to the standard deviation.) But I prefer to regard this as suggesting that the range has additional value as an indicator of the quality of the bearing. More light could be thrown on this point by an analysis of the individual snap observations, but it is of little operational importance.

[5] Provided, as will usually be the case, that the quantity $N\phi$ introduced by Beale [1960] is small.

6. Estimating Variance Components From Check Bearing Data

The fundamental statistical problem in this work can be expressed as follows:

Given n independent observations, each from a normal population with zero mean, such that the ith observation x_i has a variance

$$B_i=\sum_{j=1}^{p}\alpha_{ij}v_j, \tag{6.1}$$

where the α_{ij} are all known but some at least of the v_j are unknown, estimate the unknown v_j.

Before showing how to solve this problem, I should perhaps indicate its relevance. The quantity x_i represents the actual error in the ith bearing observation, B_i represents its assumed variance, and the v_j represent the parameters in the classification system. For example, suppose that we have a simple model with one component that can be assumed known, and unknown components depending on the radiofrequency and quality, assumed to depend on 4 parameters, such that

$$V_f=v_1+fv_2,$$

where f denotes frequency band, either 1, 2 or 3, and

$$V_q=v_3 \text{ if the swing } <8°,$$

$$V_q=v_4 \text{ if the swing } \geq 8°.$$

Now suppose that the first few check bearings have the following characteristics:

Bearing No.	*Known Variance Component*	*Frequency Band*	*Swing*
1	6	1	3°,
2	3	2	13°,
3	5	3	7°,

etc.

Then, if we write formally $v_5=1$, (6.1) is satisfied with

$$\alpha_{11}=1,\ \alpha_{12}=1,\ \alpha_{13}=1,\ \alpha_{14}=0,\ \alpha_{15}=6,$$

$$\alpha_{21}=1,\ \alpha_{22}=2,\ \alpha_{23}=0,\ \alpha_{24}=1,\ \alpha_{25}=3,$$

$$\alpha_{31}=1,\ \alpha_{32}=3,\ \alpha_{33}=1,\ \alpha_{34}=0,\ \alpha_{35}=5,$$

. . .,

the parameters to be estimated being v_1, v_2, v_3, and v_4.

We assume that our bearing errors are measured ni degrees. Variances are therefore measured in degrees squared.

Now B_1 is intended to represent the mean value of x_1^2, so it seems natural to try to minimize the sum of squares of the deviations of the x_i^2 from the corresponding B_i. But we must remember that the expected magnitudes of these deviations depend on the variances of the x_i^2. In fact if x_i is normally distributed the variance of x_i^2 is $2B_i^2$. It therefore seems reasonable to try to minimize

$$\sum_{i=1}^{n}w_i(x_i^2-B_i)^2, \tag{6.2}$$

where the weights w_i are regarded as constants but where their numerical values are given by

$$w_i=1/B_i^2. \tag{6.3}$$

Differentiating (6.2) with respect to v_j, regarding w_i as constant, we have, using (6.1)

$$\sum_{i=1}^{n}w_i\alpha_{ij}(x_i^2-B_i)=0, \tag{6.4}$$

for each j such that v_j is to be estimated.

In particular if for some j, α_{ij} is always either 0 or 1, the corresponding equation from (6.4) becomes

$$\Sigma_i' w_i(x_i^2-B_i)=0,$$

where Σ_i' denotes summation over all observations such that $\alpha_{ij}=1$.

The problem is therefore to find estimates $\hat{v}_j$ for the unknown v_j satisfying (6.1), (6.3) and (6.4). This will have to be done iteratively. Choose trial estimates for the v_j, compute the corresponding values of B_i from (6.1), and hence compute the weights w_i from (6.3), substitute these values of w_i in (6.4) and hence estimate new values of the v_j.[6] The whole cycle can be repeated if necessary, but it should be remembered that small (relative) errors in the w_i will be of little significance.

It can easily be shown that these estimates are in fact maximum likelihood estimates for the v_j. The above derivation is less rigorous but seems more intuitive.[7]

One important proviso should be made. It is undesirable to have weights for individual bearings of the same order of magnitude as the sum of the weights for all other bearings put together. In the analyses for the Brooke system, this is avoided by putting $w_i=\frac{1}{2}$ whenever $B_i<\sqrt{2}$. So the used version of (6.3) reads

$$w_i=\min\ (\tfrac{1}{2},\ 1/B_i^2).$$

It is of interest to consider the probable accuracy of the final variance estimates. It is easy to find the variance of an individual component, say v_k, if we assume that all the other components are known exactly, and also that each w_i equals the reciprocal of the square of the true variance of the

[6] In practice it may be more convenient to improve the estimates of the parameters referring to one variance component, say V_f, assuming that the others have their present trial values. This reduces the number of equations that have to be solved simultaneously.

[7] Some statisticians may point out that for this problem the method of maximum likelihood gives an "inadmissible" estimator, since it may produce negative variance components. With a fair-sized sample this is unlikely to happen; but, if it does, one must replace the negative component by some more plausible value and re-estimate the other components accordingly.

corresponding x_i. For (6.4) then becomes

$$\sum_{i=1}^{n} w_i\alpha_{ik}\left(x_i^2-\sum_{j=1}^{p}\alpha_{ij}v_j\right)=0,$$

so that the estimate $\hat{v}_k$ of v_k is given by

$$\hat{v}_k=\sum_{i=1}^{n} w_i\alpha_{ik}(x_i^2-\sum_{j\neq k}\alpha_{ij}v_j)/\Sigma w_i\alpha_{ik}^2.$$

Now var $x_i^2=2B_i^2$, and $w_i=1/B_i^2$, so

$$\text{var } \hat{v}_k=\sum_{i=1}^{n}\left(\frac{\alpha_{ik}}{B_i^2}\right)^2\cdot 2B_i^2\Big/\left(\sum_{i=1}^{n}\frac{\alpha_{ik}^2}{B_i^2}\right)^2$$

$$=2\Big/\sum\left(\frac{\alpha_{ik}^2}{B_i^2}\right)$$

$$=2\Big/\sum w_i\alpha_{ik}^2. \qquad (6.5)$$

If $\alpha_{ik}=0$ or 1 for all i, (6.5) reduces to

$$\text{var } \hat{r}_k=2/\Sigma_k' w_i, \qquad (6.6)$$

where Σ_k' denotes summation over those check bearings for which $\alpha_{ik}=1$, i.e., for which the given variance component is relevant.

Errors in estimating the weights w_i, and also errors in the other variance components, will not normally have a major effect on the variance of $\hat{v}_k$. On the other hand the fact that the observations may not be strictly independent, and that the error distribution may not be strictly normal, may have a considerable effect on the validity of these formulas. Our experience suggests that, presumably because of these causes, formulas (6.5) and (6.6) underestimate the variances of the $\hat{v}_k$ by a factor of about 2. The formulas

$$\text{var } \hat{v}_k\approx 4/\Sigma w_i\alpha_{ik}^2, \text{ or var } \hat{v}_k\approx 4/\Sigma_k' w_i,$$

are therefore recommended in practice.

There is one pecularity about these variance component estimates that one should be aware of, although it does not affect the operation of the system in any way. There is a degree of indeterminacy about the system in that a constant can be subtracted from all values of, say, V_d and added to all values of, say, V_f without altering any of the Brooke variances. The only restriction is that none of the variance components can ever be negative. Our practice is to put the minimum values of V_d and V_q equal to zero, and to put the "unattached variance" into V_f. This is purely for operational convenience. In fact we believe that the minimum value of the effect of lateral deviation, which is represented by V_d, is about 1° squared, as indicated in section 5 above.

We have now considered the estimation of variance components in the first instance.

The other vital element in an objective bearing classification system is a method of keeping the system up-to-date. For routine corrections to be timely and not unduly subject to sampling fluctuations, the corrections should be based on as simple a statistical model as possible. We therefore assume that all variances of bearings taken at a particular station are underestimated by a constant small amount. The best estimate of this amount is then approximately

$$\Sigma w_i\ (x_i^2-B_i)/\Sigma w_i,$$

where summation extends over all bearings taken at this station. This quantity is called the "apparent correction", short for "apparent Brooke variance correction", and its variance is approximately $4/\Sigma w_i$. If we wait until $\Sigma w_i\geq 20$, the variance of the apparent correction is less than about 0.2, so the correction is unlikely to be in error by as much as one unit. Nevertheless the apparent correction is rounded down in absolute value to the nearest unit, since some weight should be given to the fact that—if the system has been running for some time—previous data suggested that no correction was necessary.

In practice this works as follows. For each check bearing the "weight" w_i and the "indicator" $w_i(x_i^2-B_i)$ are recorded, and these are summed for all check bearings taken in a month at each station. These data are inspected and combined with those for enough previous months to make $\Sigma w_i\geq 20$ for the station in question. Then $\Sigma w_i(x_i^2-B_i)/\Sigma w_i$, rounded down in absolute value to the nearest unit, is added to all values of V_f for this station; i.e., to the value of a_f in the formula $V_f=a_f+b_f f$.

After any adjustment to the value of a_f, the monthly totals of $\Sigma w_i(x_i^2-B_i)$ must obviously be reduced by Σw_i multiplied by the addition to a_f before they are combined with subsequent monthly totals to decide whether further changes should be made. But it should never be necessary to recompute the B_i and w_i for individual bearings.

In practice our estimated variance components are not subject to frequent oscillations.

In addition to these routine adjustments to the overall variances for each station, it is important that the other variance components should be reviewed from time to time, by computing apparent corrections from the formula $\Sigma w_i(x_i^2-B_i)/\Sigma w_i$ with summation extending over all check bearings associated with a particular value of the component under review. The desirability of introducing further terms into the expression for the variance can be tested in the same way.

This may be as good a place as any to mention the problems of systematic errors and wild bearings. Many DF stations exhibit small systematic errors in their bearings. These have been studied for some time, but corrections to allow for them are not applied because the physical basis of the errors is not understood. If the systematic error is ignored, one is acting as if the error distribution had mean zero.

and the Brooke variance is really an estimate of the mean square error.

DF stations are apt to produce a certain number of wild bearings that are a long way off the true bearing, and which one hopes to be able to reject in plotting by their inconsistency with the other bearings. Obviously one does not want to include these in one's check bearing data for the determination of variances. But the problem is, how large must the error be before one is justified in rejecting the bearing as wild on the grounds that it would probably be recognized as such even when the true target position was unknown? This is not an easy problem. The proper answer presumably depends on the variance of nonwild observations, and on the number, location, and accuracy of the other DF stations in the organization working on the same tasks. The rejection of bearing errors greater than 10 or 15° may often be reasonable.

Although it has fallen to me to present this account of the Brooke system, I hope I have made it clear that most of the credit belongs elsewhere. I have been able to refer to some of the relevant work by Messrs. Ross, Brooke, and Bowen; but many people have contributed in various ways—notably those who organize the check bearing program, and collate and analyze the data.

7. References

Beale, E. M. L., Confidence regions in nonlinear estimation, J. Roy. Stat. Soc. (B) **22,** 41–88 (1960).

Beale, E. M. L., The estimation of variances of position lines from fixes with unknown target positions, J. Research NBS **65D** (Radio Prop.) (May-June 1961).

Bowen, K. C., Sources of error in U-Adcock high-frequency direction-finding, Proc. I.E.E. **102** Pt. B, 529–532 (1955).

Ross, W., The estimation of the probable accuracy of high frequency radio direction-finding bearings, Journal I.E.E. **94** Pt. 3A, 722–726 (1947).

Ross, W., Lateral deviation of radio waves reflected at the ionosphere, DSIR Radio Research Special Rept. No. 19 (1949).

Tippett, L. H. C., On the extreme individuals and the range of samples taken from a normal population, Biometika **17** 364–387 (1925).

(Paper 65D3—128)

JOURNAL OF RESEARCH of the National Bureau of Standards—D. Radio Propagation
Vol. 65D, No. 3, May–June 1961

Estimation of Variances of Position Lines From Fixes With Unknown Target Positions [1, 2]

E. M. L. Beale

(August 22, 1960; revised December 27, 1960)

Formulas are derived for the estimation of the variances of position lines from fixes with unknown target positions. Two approaches are considered, (1) that presented by Daniels [1951], and (2) an analysis of the squares of the errors in the position lines assuming the target is at the least squares estimate for its position.

1. Introduction

An important problem in position fixing is to estimate the accuracy of one's position lines. This can be done by assigning a variance to each line, see for example Daniels [1951] and Beale [1961]. These variances are usually estimated from fixes on targets whose true positions are known to the analyst. There are two major difficulties about estimating variances from fixes where the true position is unknown: (1) The estimates depend critically on the assumption that errors in different position lines are independent; and (2) The statistical problem is difficult, and any valid method seems to involve a substantial computing effort.

In spite of these difficulties, the problem deserves attention, because it is sometimes impossible to set up a satisfactory program of check fixes on known targets. Two approaches are considered in this paper. One follows Daniels [1951], the other is based on an analysis of the squares of the errors in the position lines assuming that the target is at the least squares estimate for its position.

Section 2 of this paper presents the basic assumptions common to both these approaches. The next 3 sections are concerned with Professor Daniels' approach to the problem. The approach is presented in general terms in section 3. Specific formulas for the important special case of 4 lines per fix are developed in section 4; and the application of these formulas is considered in section 5. The formulas for the alternative approach are developed in section 6, and their application is considered in section 7.

Finally, in section 8 some suggestions are made for an artificial sampling experiment to try out both these approaches.

2. Basic Assumptions

The basic assumptions made in this work are: (1) The earth is flat near the true position; (2) the position lines are straight lines; (3) an error of observation displaces the line parallel to itself; (4) the errors have zero means and are statistically independent; and (5) the errors are normally distributed, and we have rough estimates of their variances.

Given these assumptions, we can take Cartesian coordinates, and denote the j^{th} position line by the equation

$$x \sin \theta_j - y \cos \theta_j = p_j, \tag{2.1}$$

where θ_j is a known constant, and p_j is regarded as a random variable with mean

$$\xi \sin \theta_j - \eta \cos \theta_j,$$

[1] Contribution from Admiralty Research Laboratory, Teddington, Middlesex, England.
[2] Paper presented at the Conference on Transmission Problems Related to High-Frequency Direction Finding, at UCLA, June 21–24, 1960.

and variance σ_j^2, where (ξ,η) are the unknown coordinates of the true position. In our theoretical work, we assume that $\xi=\eta=0$, so that the mean value of p_j is zero; but we must remember that p_j is then not a directly observable quantity.

Some discussion of these basic assumptions may be helpful.

The first 3 assumptions are introduced to make the model linear, i.e., to make the value of each observation in the absence of errors a linear function of the unknown parameters ξ and η. It is clear that in some DF situations, notably when the DF stations have a narrow base line, the problem is decidedly nonlinear when expressed in terms of obvious parameters, such as the distances of the target east and north of some origin. But there is evidence to suggest that the "intrinsic nonlinearity" as defined by Beale [1960], is nearly always small for DF problems. This implies that any results derived from the standard linear theory will be valid, provided that they are expressed in terms independent of any particular system of parameters (and provided of course that any other assumptions made about the error distributions are valid).

Using the language of Beale [1960], we must define all the quantities we use in terms of sample space and the solution locus, and not solely in terms of parameter space (or possible target positions on the Earth, which is a representation of parameter space). Now a position line can be defined in sample space as the intersection with the solution locus of the hyperplane where one coordinate, i.e., one observation, is constant. All we require in this work is an interpretation of the quantities p_j, θ_j, and σ_j^2. This can be obtained by taking some specific point T on the Earth near the true target position, and imagining a parameter system that coincides locally with distances east and north of T. Now let β_j denote the actual j^{th} observation. (In the DF problem this is the bearing from the j^{th} DF station.) Let β_{jT} denote a hypothetical j^{th} observation giving a position line passing through T, and let $\rho_j(\beta_j)$ denote the (signed) distance on the Earth of the j^{th} position line from T as a function of β_j.

Let λ_j denote $d\ \rho_j/d\beta_j$, evaluated where $\beta_j=\beta_{jT}$.

Then we can write $p_j=\lambda_j(\beta_j-\beta_{jT})$, $\sigma_j^2=\lambda_j^2\text{var}\ (\beta_i)$, and θ_j is the angle that the hypothetical j^{th} position line passing through T makes with some arbitrary x-axis passing through T.

Note that it is not necessarily legitimate to interpret θ_j as the angle that the observed position line on the Earth makes with some x-axis; or to interpret λ_j as $d\rho_j/d\beta_j$ evaluated for the observed β_j. It is important that the same point T should be used for all position lines in the fix. (This approach can be used to derive the standard (Gauss) iterative procedure for finding the least squares estimate for the target position when the variances of the β_j are assumed known. The point T must then be taken as the trial value of the least squares point at each stage.)

Our 4th basic assumption, that the error distributions have zero means and are independent, is very important. In certain circumstances it will not be satisfied in practice, in which case the least squares point will usually be a less accurate estimate of the target's true position than standard theory suggests, even if the variances are correctly estimated; and furthermore the variances will be underestimated by the methods described in this paper. If we had numerical values for all correlations involved, we could allow for them, but in general there seems to be no practical alternative to assuming independence, hoping for the best, but realizing the possibility of being misled by correlated data.

Our 5th assumption, that the errors are normally distributed and we have rough estimates of their variances, is not so critical. It is only required to give appropriate weights in our least squares estimation of the variances, and to estimate the accuracy of our final variance estimates.

3. Daniels' Approach in General

The problem considered in this paper was considered by Daniels [1951]. In section 9, Daniels writes:

"When the true position corresponding to each cocked hat is unknown, . . . it is found that the method of maximum likelihood is useless for estimating the variances of the lines. It gives in general inconsistent answers, and sometimes no answer at all . . . . The method [breaks down because] the number of incidental parameters to be estimated increases with the number of cocked hats used.

"Nevertheless a simple method of estimating the variances does exist requiring no assumptions about the error distributions, provided parallel displacement is assumed and the errors are independent. Suppose there are N three-line cocked hats, not necessarily for the same true position, the lines having variances σ_1^2, σ_2^2, σ_3^2 to be determined. The expression

$$u=p_1 \sin(\theta_2-\theta_3)+p_2 \sin(\theta_3-\theta_1)+p_3 \sin(\theta_1-\theta_2) \tag{3.1}$$

has the same value whatever the point from which the perpendiculars p_1, p_2, and p_3 are dropped on to the lines, and u is easily measured by taking the point at a vertex. . . . If the point is chosen at the unknown true position it follows that

$$E(u^2)=\sigma_1^2 \sin^2(\theta_2-\theta_3)+\sigma_2^2 \sin^2(\theta_3-\theta_1)+\sigma_3^2 \sin^2(\theta_1-\theta_2). \tag{3.2}$$

If $N \geq 3$, the N equations with the observed values of u^2 replacing $E(u^2)$ may be solved by least squares. Since for normal errors var $(u^2)=2\{E(u^2)\}^2$, arbitrary weights have to be used in the first solution and correct weights approximated to in subsequent iterations. The estimates are, however, unbiased for any error distribution whatever the weights used."

But it should be noted that with fixed DF stations the normal equations derived from (3.2) to estimate the variances are liable to be very ill conditioned, and may even be singular. If the simplifying assumption about parallel displacements were strictly correct, then the values of θ_1, θ_2, and θ_3 would be the same for all tasks, so that the right hand sides of all equations of the form (3.2) would be strictly proportional, and one could never hope to estimate more than the given linear function of σ_1^2, σ_2^2, and σ_3^2. One might hope in practice to be saved by the error in the approximation; since θ_j may be effectively constant for all reasonably possible observations on a given target, but not constant over all targets. But this will only be so if the targets are well distributed about the DF stations.

It is fairly easy to see that this difficulty is fundamental to the problem, and is not simply a defect in the present approach. For if the θ_j were strictly constant, one could for example assume that $\sigma_1=\sigma_2=0$, and the signed distance of the point of intersection of the first two position lines from the third will have a certain probability distribution (normal if the bearing errors are normal) with mean zero and variance given by (3.2). This variance is constant as long as the true values of σ_1^2, σ_2^2, and σ_3^2 are constant. So data of this type can never disprove the hypothesis that $\sigma_1=\sigma_2=0$.

In the analogous one-dimensional problem, if instruments 1 and 2 give independent observations x_1 and x_2 of an unknown scalar quantity ξ, and x_i has mean ξ and variance σ_i^2, then $E(x_1-x_2)^2=\sigma_1^2+\sigma_2^2$. With an arbitrarily large number of pairs of observations one can therefore get an arbitrarily accurate estimate of $\sigma_1^2+\sigma_2^2$, but cannot estimate σ_1^2 or σ_2^2 separately. But if a third independent instrument is available, then one can estimate $\sigma_1^2+\sigma_2^2$ by $(x_1-x_2)^2$, $\sigma_1^2+\sigma_3^2$ by $(x_1-x_3)^2$, and $\sigma_2^2+\sigma_3^2$ by $(x_2-x_3)^2$. Hence σ_1^2 is estimated by the mean value of $\frac{1}{2}(x_1-x_2)^2+\frac{1}{2}(x_1-x_3)^2-\frac{1}{2}(x_2-x_3)^2$, i.e., $x_1^2-x_1x_2-x_1x_3+x_2x_3$, as pointed out by Pearson [1902].

If one adds a fourth position line to the 2-dimensional problem, then one can obtain four expressions analogous to (3.2) by taking each set of three lines in turn. It turns out that these four expressions do not suffice to estimate the four variances σ_1^2, σ_2^2, σ_3^2, and σ_4^2; because the determinant of the left hand side of the four equations vanishes identically. But Professor Daniels has pointed out in a private communication that one can get more expressions by considering the mean values of the products of u-statistics for different triangles. Thus if

$$u_{123}=p_1 \sin(\theta_2-\theta_3)+p_2 \sin(\theta_3-\theta_1)+p_3 \sin(\theta_1-\theta_2),$$

and

$$u_{124}=p_1 \sin(\theta_2-\theta_4)+p_2 \sin(\theta_4-\theta_1)+p_4 \sin(\theta_1-\theta_2),$$

then

$$E(u_{123}u_{124})=\sigma_1^2 \sin(\theta_2-\theta_3) \sin(\theta_2-\theta_4)+\sigma_2^2 \sin(\theta_3-\theta_1) \sin(\theta_4-\theta_1).$$

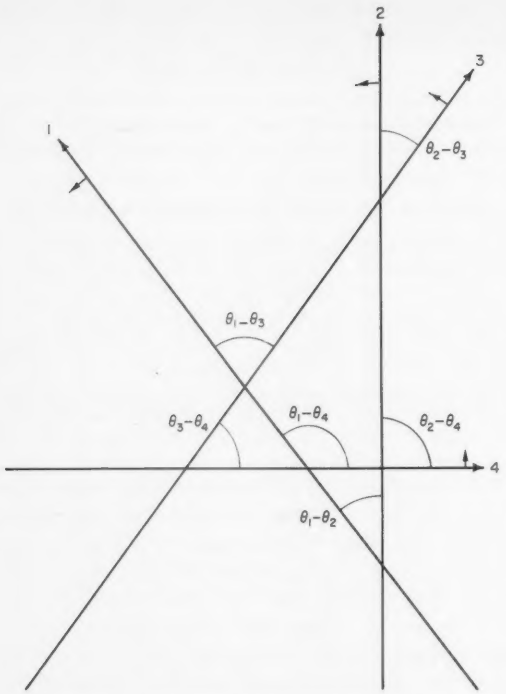


FIGURE 1. *A possible set of 4 position lines.*

Attention must be paid to signs. It is convenient to regard each position line as a directed line, pointing in the direction θ from some apex at an arbitrarily large distance from the scene of action. One can then define the corresponding displacement p as positive for points whose bearing from the apex is greater than θ (by an arbitrarily small amount) and negative on the other side of the position line. This will ensure that the signs of the terms on the right hand sides of the expressions analogous to (3.1) come out as expected. But it will still be necessary to study the configuration of lines to see which u-statistics are positive and which are negative. In the situation illustrated in figure 1, u_{134} and u_{234} are positive, and u_{123} and u_{124} are negative.

Given fixes with $n>3$ position lines per fix, one can obtain $U=n(n\text{-}1)(n\text{-}2)/6$ u-statistics from each fix, and $U(U+1)/2$ derived statistics by considering the squares and products of the u-statistics.

If $n=4$, this gives 4 u-statistics and 10 derived statistics.

If $n=5$, we have 10 u-statistics and 55 derived statistics.

This approach is therefore not very practical in its present form for $n\geq 5$. But the case $n=4$ is important, as it is the smallest n for which unique variance estimates can be obtained from a set of fixes, each with essentially the same values of θ_j, though not all on the same target. We therefore explore this case in more detail in the next section.

4. Daniels' Approach With $n=4$

With 4 position lines one can form 4 triangles, and obtain 4 basic quantities u_{123}, u_{124}, u_{134}, and u_{234}, which have the following expressions

$$\left.\begin{aligned} u_{123} &= p_1 s_{23} + p_2 s_{31} + p_3 s_{12} \\ u_{124} &= p_1 s_{24} + p_2 s_{41} + p_4 s_{12} \\ u_{134} &= p_1 s_{34} + p_3 s_{41} + p_4 s_{13} \\ u_{234} &= p_2 s_{34} + p_3 s_{42} + p_4 s_{23}, \end{aligned}\right\} \tag{4.1}$$

where s_{ij} denotes sin $(\theta_i - \theta_j)$.

From these one can derive a vector $\underset{\sim}{y}$ of 10 derived observations, such that

$$\underset{\sim}{y} = \underset{\sim}{A}\underset{\sim}{q}, \tag{4.2}$$

where

$$\left.\begin{aligned} &y_1 = u_{123}^2,\ y_2 = u_{123}u_{124},\ y_3 = u_{123}u_{134},\ y_4 = u_{123}u_{234}, \\ &y_5 = u_{124}^2,\ y_6 = u_{124}u_{134},\ y_7 = u_{124}u_{234}, \\ &y_8 = u_{134}^2,\ y_9 = u_{134}u_{234}, \\ &y_{10} = u_{234}^2; \end{aligned}\right\} \tag{4.3}$$

$$\left.\begin{aligned} &q_1 = p_1^2,\ q_2 = p_1 p_2,\ q_3 = p_1 p_3,\ q_4 = p_1 p_4, \\ &q_5 = p_2^2,\ q_6 = p_2 p_3,\ q_7 = p_2 p_4, \\ &q_8 = p_3^2,\ q_9 = p_3 p_4, \\ &q_{10} = p_4^2; \end{aligned}\right\} \tag{4.4}$$

and A is

$$\begin{bmatrix} s_{23}^2 & -2s_{13}s_{23} & 2s_{12}s_{23} & 0 & s_{13}^2 & -2s_{12}s_{13} & 0 & s_{12}^2 & 0 & 0 \\ s_{23}s_{24} & (-t_2-t_3) & s_{12}s_{24} & s_{12}s_{23} & s_{13}s_{14} & -s_{12}s_{14} & -s_{12}s_{13} & 0 & s^2{}_{12} & 0 \\ s_{23}s_{34} & -s_{13}s_{34} & (t_1-t_3) & s_{13}s_{23} & 0 & s_{13}s_{14} & -s_{13}^2 & -s_{12}s_{14} & s_{12}s_{13} & 0 \\ 0 & s_{23}s_{34} & -s_{23}s_{24} & s_{23}^2 & -s_{13}s_{34} & (t_1+t_2) & -s_{13}s_{23} & -s_{12}s_{24} & s_{12}s_{23} & 0 \\ s_{24}^2 & -2s_{14}s_{24} & 0 & 2s_{12}s_{24} & s_{14}^2 & 0 & -2s_{12}s_{14} & 0 & 0 & s_{12}^2 \\ s_{24}s_{34} & -s_{14}s_{34} & -s_{14}s_{24} & (t_1+t_2)0 & & s_{14}^2 & -s_{13}s_{14} & 0 & -s_{12}s_{14} & s_{12}s_{13} \\ 0 & s_{24}s_{34} & -s_{24}^2 & s_{23}s_{24} & -s_{14}s_{34} & s_{14}s_{24} & (t_1-t_3) & 0 & -s_{12}s_{24} & s_{12}s_{23} \\ s_{34}^2 & 0 & -2s_{14}s_{34} & 2s_{13}s_{34} & 0 & 0 & 0 & s_{14}^2 & -2s_{13}s_{14} & s_{13}^2 \\ 0 & s_{34}^2 & -s_{24}s_{34} & s_{23}s_{34} & 0 & -s_{14}s_{34} & s_{13}s_{34} & s_{14}s_{24} & (-t_2-t_3)s_{13}s_{23} & \\ 0 & 0 & 0 & 0 & s_{34}^2 & -2s_{24}s_{34} & 2s_{23}s_{34} & s_{24}^2 & -2s_{23}s_{24} & s_{23}^2 \end{bmatrix} \tag{4.5}$$

where $t_1 = s_{12}s_{34}$, $t_2 = s_{13}s_{24}$ and $t_3 = s_{14}s_{23}$.

Then

$$E(\underset{\sim}{y}) = \underset{\sim}{H}(\sigma_1^2,\ \sigma_2^2,\ \sigma_3^2,\ \sigma_4^2)', \tag{4.6}$$

where $\underset{\sim}{H}$ is the matrix formed from the 1st, 5th, 8th, and 10th columns of $\underset{\sim}{A}$, since the p_j are independent and have zero means when referred to the true target position as origin.

But in order to determine rational, as opposed to arbitrary, least squares estimates for the unknown parameters σ_1^2, σ_2^2, σ_3^2, and σ_4^2, it is necessary to consider the covariance matrix of $\underset{\sim}{y}$, We rely on the fact that the p_j are independent and have zero means, and find that

$$c_{ij} = \operatorname{cov}(y_i y_j) = Ey_i y_j - (Ey_i)(Ey_j) = \sum_k a_{ik}a_{jk}r_k, \tag{4.7}$$

where

$$\left.\begin{aligned} &r_1 = 2\sigma_1^4,\ r_2 = \sigma_1^2\sigma_2^2,\ r_3 = \sigma_1^2\sigma_3^2,\ r_4 = \sigma_1^2\sigma_4^2, \\ &r_5 = 2\sigma_2^4,\ r_6 = \sigma_2^2\sigma_3^2,\ r_2 = \sigma_2^2\sigma_4^2, \\ &r_8 = 2\sigma_3^4,\ r_9 = \sigma_3^2\sigma_4^2, \\ &r_{10} = 2\sigma_4^4. \end{aligned}\right\} \tag{4.8}$$

(The coefficients 2 in r_1, r_5, r_8, and r_{10} should be replaced by $2+\kappa_4$ if the error distributions have a fourth cumulant κ_4 different from zero.)

The covariance matrix $\underset{\sim}{C}=(c_{ij})$ of the derived observations is therefore obtained from (4.7), or equivalently by defining a new matrix

$$\underset{\sim}{F}=(f_{ij}),$$

$$\text{where } f_{ij}=a_{ij}r_j^{1/2},$$

$$\text{so that } \underset{\sim}{C}=\underset{\sim}{F}\,\underset{\sim}{F}'.$$

Now the standard least squares results can be expressed in matrix notation as follows: If $\underset{\sim}{Y}=\underset{\sim}{X}\underset{\sim}{\beta}+\underset{\sim}{\epsilon}$, where $\underset{\sim}{Y}$ is a vector of observations, $\underset{\sim}{X}$ is a known matrix, $\underset{\sim}{\beta}$ is a vector of unknown parameters, and $\underset{\sim}{\epsilon}$ is a random vector whose components are uncorrelated and have means zero and a common variance σ^2, then the least squares estimates $\underset{\sim}{B}$ of $\underset{\sim}{\beta}$ are given by choosing $\underset{\sim}{B}$ to minimize

$$Q=(\underset{\sim}{X}\underset{\sim}{B}-\underset{\sim}{Y})'(\underset{\sim}{X}\underset{\sim}{B}-\underset{\sim}{Y}), \tag{4.9}$$

Assuming that the errors are normally distributed, the likelihood is proportional to $\exp(-Q/2\sigma^2)$, and in any case

$$\underset{\sim}{B}=(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Y}, \tag{4.10}$$

and the covariance matrix for the estimates $\underset{\sim}{B}$ is given by $(\underset{\sim}{X}'\,\underset{\sim}{X})^{-1}\sigma^2$. (4.11)

If now $\underset{\sim}{y}=\underset{\sim}{G}\underset{\sim}{Y}$, we have $\underset{\sim}{y}=\underset{\sim}{G}\underset{\sim}{X}\underset{\sim}{\beta}+\underset{\sim}{G}\underset{\sim}{\epsilon}$, and therefore the covariance of y_i and y_j is given by

$$\sum_k g_{ik}g_{jk}\sigma^2.$$

Hence we can identify $\underset{\sim}{G}$ with the matrix $\underset{\sim}{F}$, putting $\sigma^2=1$, and we know that, if $\beta=(\sigma_1^2,\ \sigma_2^2,\ \sigma_3^2,\ \sigma_4^2)'$,

$$\underset{\sim}{y}=\underset{\sim}{H}\underset{\sim}{\beta}+\underset{\sim}{F}\underset{\sim}{\epsilon}.$$

So $\underset{\sim}{X}=\underset{\sim}{F}^{-1}\underset{\sim}{H}$, $\underset{\sim}{Y}=\underset{\sim}{F}^{-1}\underset{\sim}{y}$,

and

$$\begin{aligned}Q&=\{\underset{\sim}{F}^{-1}(\underset{\sim}{H}\underset{\sim}{\beta}-\underset{\sim}{y})\}'\{\underset{\sim}{F}^{-1}(\underset{\sim}{H}\underset{\sim}{\beta}-\underset{\sim}{y})\}\\&=(\underset{\sim}{H}\underset{\sim}{\beta}-\underset{\sim}{y})'\underset{\sim}{F}^{-1\prime}\underset{\sim}{F}^{-1}(\underset{\sim}{H}\underset{\sim}{\beta}-\underset{\sim}{y})\\&=(\underset{\sim}{H}\underset{\sim}{\beta}-\underset{\sim}{y})'(\underset{\sim}{F}\underset{\sim}{F}')^{-1}(\underset{\sim}{H}\underset{\sim}{\beta}-\underset{\sim}{y})\\&=(\underset{\sim}{H}\underset{\sim}{\beta}-\underset{\sim}{y})'\underset{\sim}{C}^{-1}(\underset{\sim}{H}\underset{\sim}{\beta}-\underset{\sim}{y}).\end{aligned} \tag{4.12}$$

Further

$$\begin{aligned}\underset{\sim}{\beta}&=(\underset{\sim}{H}'\underset{\sim}{F}'^{-1}\underset{\sim}{F}^{-1}\underset{\sim}{H})^{-1}\underset{\sim}{H}'\underset{\sim}{F}'^{-1}\underset{\sim}{F}^{-1}\underset{\sim}{y}\\&=(\underset{\sim}{H}'\underset{\sim}{C}^{-1}\underset{\sim}{H})^{-1}\underset{\sim}{H}'\underset{\sim}{C}^{-1}\underset{\sim}{y},\end{aligned} \tag{4.13}$$

and the covariance matrix of $\underset{\sim}{\beta}$ is given by

$$(\underset{\sim}{H}'\underset{\sim}{C}^{-1}\underset{\sim}{H})^{-1}. \tag{4.14}$$

These results are due to Aitken [1935].

5. Application of the Formulas of Section 4

To apply these formulas to a situation in which one has N fixes, i.e., sets of four position lines, each with a constant variance and a constant value of θ_j, one proceeds as follows.

One first sorts out one's sign convention, as indicated at the end of section 3. This will produce signed values of u_{123}, u_{124}, u_{134}, and u_{234} for each of the N fixes. From these one computes derived observations $y_1, \ldots, y_{10}$ for each fix from (4.3). Averaging over all fixes, one obtains a vector

$$\underset{\sim}{\bar{y}}=(\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_{10})'.$$

We next compute the matrix $\underset{\sim}{A}$ from (4.5).

By using the guessed values for the unknown variances σ_1^2, σ_2^2, σ_3^2, and σ_4^2, one can derive values of $r_1^2, r_2^2, \ldots, r_{10}^2$ from (4.8), and hence compute the elements c_{ij} of the matrix $\underset{\sim}{C}$ from (4.7).

This matrix must be inverted to produce $\underset{\sim}{C}^{-1}$.

We must also consider the matrix $\underset{\sim}{H}$ whose columns are the 1st, 5th, 8th, and 10th columns of $\underset{\sim}{A}$, and our estimates of σ_1^2, σ_2^2, σ_3^2, and σ_4^2 are the 4 elements of the matrix

$$(\underset{\sim}{H}'\underset{\sim}{C}^{-1}\underset{\sim}{H})^{-1}\underset{\sim}{H}'\underset{\sim}{C}^{-1}\underset{\sim}{\bar{y}}. \tag{5.1}$$

If our guessed values of the σ_i^2, used to define $\underset{\sim}{r}$, were correct, the covariance matrix for these estimates would be

$$\frac{1}{N}(\underset{\sim}{H}'\underset{\sim}{C}^{-1}\underset{\sim}{H})^{-1}. \tag{5.2}$$

One could try iterating this procedure, using the estimated variances to produce a new vector $\underset{\sim}{r}$, and hence a new $\underset{\sim}{C}$, and hence revised estimated variances, and so on.

I am doubtful of the wisdom of this, though it might be interesting to try it on an artificial sampling experiment. If one wants a reasonable estimate of the accuracy of the estimates one must compute a revised $\underset{\sim}{C}$ to use in (5.2). But I can see little merit and some danger in revising the weights to produce revised estimates: Note that the iterated estimates are not unbiased.

This procedure may on occasion produce negative estimated variances. Steps must be taken at least to ensure that negative (or even very small positive) variances are not used in the definition of $\underset{\sim}{r}$. Theoretically the least squares problem should be set up as a quadratic programing problem when such negative estimates are produced by the standard method. But this is too large an issue to face at this stage.

6. A Direct Approach

In this section we derive some equations concerning the apparent error in the position lines, assuming that the target is at the least squares estimate for its position.

Suppose that the j^{th} position line is

$$x \sin \theta_j - y \cos \theta_j = p_j,$$

with the origin at the (unknown) true target position. Suppose further that we have some rough estimate s_j^2 for σ_j^2.

We write

$$\sigma_j^2 = s_j^2(1+\delta_j), \text{ and } 1/s_j^2 = w_j.$$

Now the sum of squares function, whose minimum value defines the least squares estimate, is

$$Q=\sum_j w_j(x \sin \theta_j - y \cos \theta_j - p_j)^2 = a x^2+2hxy+by^2+2gx+2fy+c,$$

where

$$\left.\begin{aligned} &a=\sum_j w_j \sin^2 \theta_j, \quad b=\sum_j w_j \cos^2 \theta_j, \quad c=\sum_j w_j p_j^2, \\ &f=\sum_j w_j p_j \cos \theta_j, \quad g=-\sum w_j p_j \sin \theta_j, \quad h=-\sum_j w_j \sin \theta_j \cos \theta_j. \end{aligned}\right\} \tag{6.1}$$

Note that a, b, and h are known constants, while f, g, and c depend on the random variables p_j.

It is convenient to rotate axes, so that in the new coordinates (X, Y),

$$Q=AX^2+BY^2+2GX+2FY+C.$$

This involves finding an angle α such that

$$\phi_j=\theta_j+\alpha, \tag{6.2}$$

and the jth position line is

$$X \sin \phi_j - Y \cos \phi_j = p_j,$$

i.e.,

$$X \sin (\theta_j+\alpha) - Y \cos (\theta_j+\alpha) = p_j,$$

so that

$$x = X \cos \alpha + Y \sin \alpha,$$

$$y = -X \sin \alpha + Y \cos \alpha.$$

And since $ax^2+2hxy+by^2+2gx+2fy+c=AX^2+2HXY+BY^2+2GX+2FY+C$, where $H=0$, we have

$$\left.\begin{aligned} A &= a \cos^2 \alpha - 2h \cos \alpha \sin \alpha + b \sin^2 \alpha, \\ 0 &= a \cos \alpha \sin \alpha + h(\cos^2 \alpha - \sin^2 \alpha) - b \cos \alpha \sin \alpha, \\ B &= a \sin^2 \alpha + 2h \cos \alpha \sin \alpha + b \cos^2 \alpha, \\ G &= g \cos \alpha - f \sin \alpha, \\ F &= g \sin \alpha + f \cos \alpha, \\ C &= c. \end{aligned}\right\} \tag{6.3}$$

This implies that

$$\tan 2\alpha = 2h/(b-a). \tag{6.4}$$

This defines a set of values of α differing by multiples of 90°. It does not matter which is taken.

The values of A, B, C, F, and G can be obtained either from (6.1) and (6.3), or alternatively from the formulas

$$\left.\begin{aligned} &A=\sum_j w_j \sin^2 \phi_j, \quad B=\sum_j w_j \cos^2 \phi_j, \quad C=\sum_j w_j p_j^2, \\ &F=\sum_j w_j p_j \cos \phi_j, \quad G=-\sum_j w_j p_j \sin \phi_j, \quad H=-\sum_j w_j \sin \phi_j \cos \phi_j = 0, \end{aligned}\right\} \tag{6.5}$$

where ϕ_j is obtained by substituting for α from (6.4) in (6.2).

The transformed coordinates of the least squares estimate are then

$$X_0=-G/A \text{ and } Y_0=-F/B. \tag{6.6}$$

Now let d_k denote the signed displacement of the k^{th} position line from the least squares estimate. This quantity can be observed in practice. Then

$$d_k = p_k + \frac{G}{A} \sin \phi_k - \frac{F}{B} \cos \phi_k = \sum_{i=1}^{n} p_i (w_i/w_k)^{1/2} \lambda_{ik}, \tag{6.7}$$

where

$$\lambda_{kk} = 1 - w_k \left(\frac{\sin^2 \phi_k}{A} + \frac{\cos^2 \phi_k}{B} \right), \tag{6.8}$$

$$\lambda_{ik} = \lambda_{ki} = -(w_i w_k)^{1/2} \left(\frac{\sin \phi_i \sin \phi_k}{A} + \frac{\cos \phi_i \cos \phi_k}{B} \right), \ (i \neq k). \tag{6.9}$$

Note that

$$\sum_{k=1}^{n} \lambda_{kk} = n - \sum w_k \sin^2 \phi_k / A - \sum w_k \cos^2 \phi_k / B = n-2. \tag{6.10}$$

Further

$$\sum_{i=1}^{n}\lambda_{ik}^2=1-2w_k\left(\frac{\sin^2\phi_k}{A}+\frac{\cos^2\phi_k}{B}\right)+w_k\sum_{i=1}^{n}\left\{\frac{w_i\sin^2\phi_i\sin^2\phi_k}{A^2}+2\frac{w_i\sin\phi_i\cos\phi_i\sin\phi_k\cos\phi_k}{AB}\right.$$
$$\left.+\frac{w_i\cos^2\phi_i\cos^2\phi_k}{B^2}\right\}=1-w_k\left(\frac{\sin^2\phi_k}{A}+\frac{\cos^2\phi_k}{B}\right).$$

So

$$\sum_{i=1}^{n}\lambda_{ik}^2=\lambda_{kk}. \tag{6.11}$$

Further, if $k\neq l$,

$$\sum_{i=1}^{n}\lambda_{ik}\lambda_{il}=2\lambda_{kl}+(w_kw_l)^{1/2}\sum_{i=1}^{n}w_i\left(\frac{\sin\phi_i\sin\phi_k}{A}+\frac{\cos\phi_i\cos\phi_k}{B}\right)\left(\frac{\sin\phi_i\sin\phi_l}{A}+\frac{\cos\phi_i\cos\phi_l}{B}\right)$$
$$=2\lambda_{kl}+(w_kw_l)^{1/2}\left(\frac{\sin\phi_k\sin\phi_l}{A}+\frac{\cos\phi_k\cos\phi_l}{B}\right)\text{ from (6.5)}=\lambda_{kl}.$$

So (6.11) can be generalized to read

$$\sum_{i=1}^{n}\lambda_{ik}\lambda_{il}=\lambda_{kl}. \tag{6.12}$$

From (6.7) we deduce that

$$w_kd_k^2=\sum_{i=1}^{n}\sum_{j=1}^{n}(w_iw_j)^{1/2}\lambda_{ik}\lambda_{jk}p_ip_j. \tag{6.13}$$

Now

$$Ep_i^2=\sigma_i^2=(1+\delta_i)/w_i,\ Ep_ip_j=0 \qquad (i\neq j),$$

So

$$Ew_kd_k^2=\sum_{i=1}^{n}\lambda_{ik}^2(1+\delta_i)=\lambda_{kk}+\sum_{i=1}^{n}\delta_i\lambda_{ik}^2, \qquad \text{from (6.11)}. \tag{6.14}$$

It follows that

$$E\sum_{k=1}^{n}w_kd_k^2=\sum_{k=1}^{n}\lambda_{kk}+\sum_{i=1}^{n}\delta_i\sum_{k=1}^{n}\lambda_{ik}^2=n-2+\sum_{i=1}^{n}\delta_i\lambda_{ii}, \qquad \text{from (6.10) and (6.11)}. \tag{6.15}$$

If $\delta_i=\delta$ for all i, we have the well-known formula

$$E\sum w_kd_k^2=(n-2)(1+\delta), \tag{6.16}$$

(which can easily be deduced from first principles).

It is also of interest to consider the covariance of $w_kd_k^2$ and $w_ld_l^2$.

We have

$$\operatorname{cov}(w_kd_k^2,\ w_ld_l^2)=\sum_{i=1}^{n}w_i^2\lambda_{ik}^2\lambda_{il}^2 2(1+\delta_i)^2/w_i^2$$

$$+2\sum_{i=1}^{n}\sum_{j\neq i}w_iw_j\lambda_{ik}\lambda_{jk}\lambda_{il}\lambda_{jl}(1+\delta_i)(1+\delta_j)/w_iw_j \qquad \text{by analogy with (4.7)}$$

$$=2\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_{ik}\lambda_{jk}\lambda_{il}\lambda_{jl}(1+\delta_i)(1+\delta_j). \tag{6.17}$$

It is of particular interest to consider the covariance matrix when all $\delta_i=0$, since this corresponds to the originally estimated covariance matrix.

Using (6.12) we find this reduces to

$$\operatorname{cov}(w_kd_k^2,\ w_ld_l^2)=2\lambda_{kl}^2. \tag{6.18}$$

If $n=3$, it can be shown that $\lambda_{11}/\lambda_{21}=\lambda_{12}/\lambda_{22}=\lambda_{13}/\lambda_{23}$, etc. For example, to prove that $\lambda_{11}\lambda_{22}-\lambda_{12}^2=0$, one substitutes for the λ_{ij} from (6.8) and (6.9) and multiplies out. The terms in $1/A^2$ and $1/B^2$ cancel, so the expression can be written as a quadratic function of w_1, w_2 and w_3 (using the definitions of A and B) divided by AB, and H proves to be a factor of the numerator.

This is an algebraic proof of something we already know on statistical grounds: that with only 3 bearings the random variables $w_1d_1^2$, $w_2d_2^2$ and $w_3d_3^2$ are totally correlated, and it is impossible to estimate more than one linear function of the unknown variances.

7. Application of the Formulas of Section 6

To apply the formulas of section 6 to a situation in which one has N fixes, i.e., sets of n position lines, each with a constant variance and a constant value of θ_j, one proceeds as follows.

One starts with some assumed variances s_j^2 for the position lines, and one uses those to compute a least squares point estimate for the target for each fix. One then computes $w_jd_j^2$, i.e., d_j^2/s_j^2, being the weighted square of the displacement of the j^{th} position line from each fix.

One then averages over all N fixes, obtaining n quantities $\overline{w_1d_1^2}$, . . ., $\overline{w_nd_n^2}$.

One must also compute λ_{ii} for $i=1, \ldots, n$, and λ_{ij}^2 for $i, j=1, \ldots, n$. These quantities are the same for all fixes.

Note that λ_{ii} is defined by (6.8) and λ_{ij} by (6.9) if $i\neq j$.

It may be more convenient to use the formula

$$\lambda_{ij}^2=\frac{(w_i\sin^2\phi_i)(w_j\sin^2\phi_j)}{A^2}+\frac{2(w_i\sin\phi_i\cos\phi_i)(w_j\sin\phi_j\cos\phi_j)}{AB}+\frac{(w_i\cos^2\phi_i)(w_j\cos^2\phi_j)}{B^2},\quad (i\neq j). \tag{7.1}$$

Then, from (6.14), we solve the equations

$$\sum_{i=1}^{n}\hat{\delta}_i\lambda_{ik}^2=\overline{w_kd_k^2}-\lambda_{kk},\ (k=1, \ldots, n). \tag{7.2}$$

and the resulting values of $\hat{\delta}_i$ define the estimated variances, since

$$\sigma_i^2\approx s_i^2(1+\hat{\delta}_i).$$

To find the estimated covariance matrix for these estimates $\hat{\delta}_i$, we form the matrix $\underset{\sim}{H}$ such that

$$(h_{ij})=\lambda_{ij}^2,$$

and the matrix $\underset{\sim}{C}$ such that

$$(c_{ij})=2\lambda_{ij}^2\quad \text{(from (6.18))}. \tag{7.3}$$

The estimated covariance matrix is then given by (5.2), if the $\hat{\delta}_i$ prove to be small, and this reduces to $\frac{4}{N}\underset{\sim}{C}^{-1}$.

If one wants to combine these data with other data, then the contribution of these data to the sum of squares to be minimized is then NQ, where Q is given by (4.12) with $\underset{\sim}{\beta}=(\delta_1, \ldots, \delta_n)'$ and $\underset{\sim}{y}=(\overline{w_1d_1^2}, \ldots, \overline{w_nd_n^2})'$.

8. Proposed Computational Program

Some computational experience with both the schemes proposed here would be very valuable.

In the first instance, it would be of interest to inspect the covariance matrix (4.14) for estimates obtained by Daniels' approach, i.e., with $\underset{\sim}{H}$ defined following (4.6) and C defined by (4.7), for various values of θ_j and σ_j^2.

It would also be of interest to inspect the corresponding matrices for the direct approach, i.e., with $(h_{ij}) = (\lambda_{ij}^2)$ and $(c_{ij}) = (2\lambda_{ij}^2)$. In this case (4.14) reduces to 4 C^{-1}. The $(ij)^{th}$ element of this covariance matrix for the $\hat{\delta}$ must be course of multiplied by $s_i^2 s_j^2$ to represent the covariance matrix for the variances. It will then be directly comparable with the other.

One may hope to throw considerable further light on these procedures by an artificial sampling experiment. One must first choose values of θ_j, σ_j^2, and (for the direct approach) s_j^2. These will define values of a, b, and h from (6.1), and hence of A, B and α from (6.3) and (6.4). One can then compute the λ_{ik} from (6.9).

One then takes sets of values of p_i as pseudo-random normal deviates with means zero and variances σ_i^2, and use these to generate synthetic u-statistics from (4.1), and values of d_i from (6.7). One can then compare the estimated variances obtained by the two approaches (a) after one iteration, and (b) after using the results of each iteration as starting values for the next (i.e., to compute the r_i for Daniels' approach and the w_i for the direct approach) and continuing until the variances that go in come out.

One can also consider the effects on the two approaches of using poor starting approximations to the variances.

It seems likely on general grounds that Daniels' approach should give better results, since it uses more information. But if the direct approach is at all satisfactory with $n=4$, one may hope that it will be still better with larger values of n.

9. References

Aitken, A. C., On least squares and linear combination of observations, Proc. Roy. Soc. Edinburgh **55,** 42 (1935).

Beale, E. M. L., Confidence regions in nonlinear estimation, J. Roy. Stat. Soc. (B) **22,** 41 (1960).

Beale, E. M. L., The Brooke variance classification system, J. Research NBS **65D**, No. 3, (May–June 1961).

Daniels, H. E., The theory of position finding, J. Roy. Stat. Soc. (B) **13,** 186 (1951).

Pearson, K., On the mathematical theory of errors of judgment, with special reference to the personal equation, Phil. Trans. Roy. Soc. London, Series *A*, **198,** 235 (1902).

(Paper 65D3–129)

JOURNAL OF RESEARCH of the National Bureau of Standards—D. Radio Propagation
Vol. 65D, No. 3, May–June 1961

Statistics of a Radio Wave Diffracted by a Random Ionosphere[1, 2, 3]

S. A. Bowhill

(August 15, 1960; revised December 27, 1960)

For some purposes, particularly in connection with the study of the random structure of the lower ionosphere, using very low frequencies, it is necessary to find the detailed statistical properties of a random signal diffracting in free space. Mathematical tools for evaluating these parameters have been developed, and are applied in this paper. Allowance is made for the effect of sphericity of the wave incident on the ionosphere, and anisotropy of the irregular variations of signal is permitted. The case of oblique incidence of a wave on the ionosphere is also considered.

1. Introduction

There has of late been an increasing interest in the irregular structure of the lower ionosphere, particularly in connection with scatter propagation. In order to interpret irregular fluctuations in radio signals observed at the ground in terms of irregularities in the ionosphere, it is first necessary to find how much the nature of the signal changes as it propagates from the ionosphere to the earth. This propagation, which takes place in free space, will be termed "diffraction," in the sense used in a previous paper [Bowhill, 1957].

Most of the methods so far proposed for dealing with this statistical problem assume that the signal at the layer is like the signal produced when an infinite plane wave passes through a thin diffracting screen. The field at the screen is then resolved into a number of plane waves propagating in various directions, their amplitudes being given by the Fourier spectrum of the field variation. The signal at any distance from the screen is then given by the vector sum of the signals from the various plane waves, combined with the appropriate phases. Booker, Ratcliffe, and Shinn [1950] stated this principle, but Hewish [1952] was the first to show quantitatively how the magnitude of the irregular field variations changed with distance from the screen. He assumed a rectangular distribution of spatial wave numbers in the screen. Feinstein [1954] has used a rather different approach based on Huygens' principle, which has also been exploited by Chernov [1960]. Ratcliffe [1956] has summarized the physical principles involved in these approaches.

Pitteway [1958] has recently given an analytic method for calculating the field due to any particular assembly of irregularities; this, however, is difficult to apply in a statistically random medium.

It is important to distinguish at this point between the various types of random screen postulated. The two types most commonly considered are screens which modulate the phase or the amplitude of the signal. A phase screen is said to be "deep" or "shallow" depending on whether the phase excursion of the signal is substantially larger or smaller than one radian. The deep phase screen problem has been approached by Hewish [1951] by using a sinusoidal variation of phase with distance, but the extension of a detailed theory to this case presents considerable difficulties. The shallow phase screen, on the other hand, is exactly equivalent to a shallow amplitude screen [Bowhill, 1959]. This interchangeability of phase and amplitude

---

[1] Contribution from Ionosphere Research Laboratory, The Pennsylvania State University, University Park, Pennsylvania.

[2] The research reported in this paper was sponsored by the Geophysics Research Directorate of the Air Force Cambridge Research Center, Air Research and Development Command, under Contract AF 19(604)-1304.

[3] Paper presented at the Conference on Transmission Problems Related to High-Frequency Direction Finding, at UCLA, June 21-24, 1960.

is exploited in the phase-contrast microscope, in which variations in phase are changed into variations in amplitude by artificially retarding the phase of the specular component by $\pi/2$. It is well known that this only gives an improvement in visibility when the phase variations are small (i.e., the phase screen is shallow).

The results of the analysis in this paper will be applied principally to the random diffraction of very low frequency radio waves. The fading for these frequencies is usually shallow indicating that a shallowly modulated random field at the ionosphere must be responsible. Attention will therefore be confined to shallowly modulated screens.

To illustrate the method of analysis, it will first be applied, in section 2, to a one-dimensional random screen. The extension to two dimensions is made in section 3. Section 4 deals with the correlation between amplitude and phase in the signal, and section 5 with the effect of wave sphericity. The work is extended to oblique incidence in section 6.

2. One-Dimensional Diffraction Problem

In treating the spatial correlation function and spatial frequency spectrum of the signal variations, it is much more convenient to use purely real quantities, rather than the complex variables of amplitude and phase. This enables the use of complex Fourier transforms to relate field variations over the screen to the complex angular spectrum of plane waves which they produce.

For an amplitude screen, which alone will be considered (in view of the equivalence indicated above), the electric field $\underline{E}$ at the screen is given by

$$\underline{E}(x_0, 0)=A(x_0)$$

where $A(x_0)$ is a purely real quantity. The field at a distance z from the screen is then defined to be

$$\underline{E}(x, z)=[E_1(x)+iE_2(x)] \exp (2\pi iz/\lambda) \tag{1}$$

where $E_1(x)$ and $E_2(x)$ are both purely real, and represent the phase components of $\underline{E}(x,z)$ in phase and in quadrature with the undiffracted portion of the incident plane wave. For a shallow screen, $E_1(x)$ gives the amplitude variation, and $E_2(x)$ the phase variation. In the analysis which follows, various statistical parameters of the quantities $E_1(x)$ and $E_2(x)$, which are actually observed in experiments at ground level, are evaluated in terms of the corresponding parameters of the original amplitude variation $A(x_0)$.

Using the notation of Booker, Ratcliffe, and Shinn [1950], we have for the angular spectrum at the screen, in terms of $s=\sin\theta$, where θ is the scattering angle,

$$P_0(s)=\int_{-\infty}^{\infty} A(x_0) \cdot \exp (-2\pi isx_0/\lambda)\, dx_0 \tag{2}$$

where λ is the wavelength. The diffracted field $\underline{E}(x,z)$ is given by the Fourier transform of this, together with the phase factor $\exp (2\pi icz/\lambda)$, where $c=\cos\theta$.

$$\begin{aligned}\underline{E}(x,z)&=\frac{1}{\lambda}\int_{-\infty}^{\infty} \exp \{2\pi i(sx+cz)/\lambda\}ds \cdot \int_{-\infty}^{\infty} A(x_0) \cdot \exp (-2\pi isx_0/\lambda)dx_0 \\ &=\frac{1}{\lambda}\iint_{-\infty}^{\infty} A(x_0) \cdot \exp \{2\pi i [(x-x_0)s+cz]/\lambda\}dx_0ds. \end{aligned} \tag{3}$$

By the definition of $E_1(x)$ in equation (1),

$$\begin{aligned}E_1(x)&=\mathscr{R}\left\{\frac{1}{\lambda}\iint_{-\infty}^{\infty} A(x_0) \cdot \exp \{2\pi i [(x-x_0)s-(1-c)z]/\lambda\}dx_0ds\right\} \\ &=\frac{1}{2\lambda}\iint_{-\infty}^{\infty} A(x_0) [\exp \{2\pi i [(x-x_0)s-(1-c)z]/\lambda\}+\exp \{-2\pi i [(x-x_0)s-(1-c)z]/\lambda\}]\, dx_0ds\end{aligned}$$

since $A(x_0)$ is purely real. The spatial spectrum $P_1(\nu)$ of $E_1(x)$ is now defined by a similar expression to (2):

$$P_1(\nu)=\int_{-\infty}^{\infty} E_1(x)\cdot\exp(-2\pi i\nu x/\lambda)dx=\frac{1}{2\lambda}\iiint_{-\infty}^{\infty} A(x_0)\cdot[\exp\{2\pi i[(x-x_0)s-\nu x-z(1-c)]/\lambda\}$$
$$+\exp\{-2\pi i[(x-x_0)s+\nu x-z(1-c)]/\lambda\}]dx_0dxds.$$

Since the integrand is a well-behaved function, and all the limits of the integration are $-\infty$ to $+\infty$, the integration can be carried out in any order. In fact, it is convenient to integrate with respect to first s and then x. With the approximation (valid for small s) that $c=1-s^2/2$, $P_1(\nu)$ is easily evaluated as

$$P_1(\nu)=\frac{1}{2}\int_{-\infty}^{\infty} A(x_0)[\exp\{2\pi i(-\nu x_0+z\nu^2/2)/\lambda\}+\exp\{2\pi i(-\nu x_0-z\nu^2/2)/\lambda\}]dx_0$$
$$=\cos(\pi z\nu^2/\lambda)\cdot\int_{-\infty}^{\infty} A(x_0)\cdot\exp(-2\pi i\nu x_0/\lambda)dx_0$$

or

$$P_1(\nu)=P_0(\nu)\cdot\cos(\pi z\nu^2/\lambda) \tag{4}$$

from equation (2). This equation gives the spatial spectrum of $E_1(x)$ at any distance z from the screen. The *power* spectrum $W_1(\nu)$ of this variation is defined by

$$W_1(\nu)=\overline{P_1(\nu)\cdot P_1^*(\nu)} \tag{5}$$

the asterisk denoting the complex conjugate, and the bar denoting averaging of the result "over systems"—i.e., the mean for a number of statistically similar screens. This quantity is essentially real, and is a relatively smoothly varying function of ν, because of the process of averaging. It can be related by equation (4) to the spatial power spectrum $W_0(\nu)$ at the screen:

$$W_1(\nu)=\overline{P_0(\nu)\cdot P_0^*(\nu)\cdot\cos^2(\pi z\nu^2/\lambda)}$$

or

$$W_1(\nu)=W_0(\nu)\cdot\cos^2(\pi z\nu^2/\lambda) \tag{6}$$

where

$$W_0(\nu)=\overline{P_0(\nu)\cdot P_0^*(\nu)}.$$

Similar quantities $P_2(\nu)$ and $W_2(\nu)$ can be defined for the quadrature component E_2 of the field (see equation (5)); a similar analysis to the preceding gives

$$\left.\begin{aligned}P_2(\nu)&=P_0(\nu)\cdot\sin(\pi z\nu^2/\lambda)\\ W_2(\nu)&=W_0(\nu)\cdot\sin^2(\pi z\nu^2/\lambda)\end{aligned}\right\} \tag{7}$$

Equations (6) and (7) enable the spatial correlograms of $E_1(x)$ and $E_2(x)$ in the diffraction pattern to be determined. Assuming that the correlogram of $A(x)$ is given, the Wiener-Khintchine theorem can be used to give $W_0(\nu)$. The power spectrum $W_1(\nu)$ is then calculated from equation (20), and the correlogram of $E_1(x)$ found by the inverse Wiener-Khintchine theorem.

This one-dimensional analysis, in which the random variations in signal are taken to occur in the x-direction only, represents an unrealistic physical model. However, as some existing work [Hewish, 1951, 1952; Jones, Millman, and Nertney, 1953] has assumed this type of variation, some results worked out for this case will be quoted in the next section, for comparison with the two-dimensional results.

3. Two-Dimensional Diffraction Problem

In this section the diffraction pattern of a two-dimensional random screen is considered, using an analysis closely similar to that employed in the previous section for the one-dimensional case. The modulation is once again assumed to be in amplitude only, as there is again an equivalence between shallow phase and amplitude screens. The effect of anisotropy of the screen will be taken into consideration by defining a spatial power spectrum $W_0(\nu,\chi)$, related to a two-dimensional spatial correlogram $P_0(\alpha,\beta)$ of the field $A(x,y)$ at the screen by a two-dimensional form of the Wiener-Khintchine thereom. As the field $A(x_0,y_0)$ is purely real, the correlogram $\rho_0(\alpha,\beta)$ is given by

$$\rho_0(\alpha,\beta)=\frac{\overline{A(x,y)\,A(x+\alpha,y+\beta)}-\overline{A(x,y)}^2}{\overline{A^2(x,y)}-\overline{A(x,y)}^2}$$

The field $\overline{A(x,y)}$, the mean field over the whole screen, has a constant phase at all points since $A(x,y)$ is real, and propagates as a single plane wave. It is convenient to imagine it removed; $A(x,y)$ then has both positive and negative values, and will be defined to have a mean square value of unity; so

$$\left.\begin{aligned}\overline{A(x_0,y_0)}&=0\\ \overline{A^2(x_0,y_0)}&=1\end{aligned}\right\} \tag{8}$$

and the correlogram is given by

$$\rho_0(\alpha,\beta)=\overline{A(x_0,y_0)A(x_0+\alpha,y_0+\beta)}.$$

The single plane wave, which will be called the "specular component" (by analogy with specular reflection from a mirror, compared with irregular reflection from a rough surface), can be reintroduced when necessary into the diffraction pattern as a constant added to $E_1(x,y)$.

Proceeding exactly as in section 2, the two phase components of the diffracted field are defined by the vector field at (x,y,z):

$$\underline{E}(x,y,z)=[E_1(x,y)+iE_2(x,y)]\,\exp\,(2\pi iz/\lambda). \tag{9}$$

The phases of the diffracted waves at z are given by purely geometrical considerations. Let the plane of one wave intersect the (x,z) plane at an angle $\sin^{-1}s$, and the (y,z) plane at an angle $\sin^{-1}r$. Then the angle between the wave front and the (x,y) plane is

$$\tan^{-1}\{(s^{-2}-1)^{-1}+(r^{-2}-1)^{-1}\}^{1/2}$$

and the phase of the wave at a distance z from the screen is

$$2\pi iz\{1+(s^{-2}-1)^{-1}+(r^{-2}-1)^{-1}\}^{-1/2}/\lambda.$$

If s and r are small, this can be expanded to give, to a first order,

$$2\pi iz(s^2/2+r^2/2)/\lambda$$

The corresponding equation to (3) for $\underline{E}(x,z)$ is therefore

$$\underline{E}(x,y,z)\,\exp\,(-2\pi iz/\lambda),$$

$$=\frac{1}{2}\iiiint_{-\infty}^{\infty} A(x_0,y_0)\cdot\exp\{2\pi i[(x-x_0)s+(y-y_0)r-s^2z/2-r^2z/2]/\lambda\}\,ds\,dr\,dx_0\,dy_0 \tag{10}$$

and by similar steps to the previous analysis, it can be shown that

$$P_1(\nu,\chi)=P_0(\nu,\chi)\cdot\cos\{\pi z(\nu^2+\chi^2)/\lambda\} \tag{11}$$

where $P_0(\nu,\chi)$ is the two-dimensional spectrum of the field at the screen. Similarly

$$P_2(\nu,\chi)=P_0(\nu,\chi)\cdot\sin\{\pi z(\nu^2+\chi^2)/\lambda\}$$

and the spatial power spectra of $E_1(x,y)$ and $E_2(x,y)$ are given by

$$\left.\begin{aligned} W_1(\nu,\chi)&=W_0(\nu,\chi).\cos^2\{\pi z(\nu^2+\chi^2)/\lambda\}\\ W_2(\nu,\chi)&=W_0(\nu,\chi).\sin^2\{\pi z(\nu^2+\chi^2)/\lambda\}.\end{aligned}\right\} \tag{12}$$

These spectra are proportional to the area of the diffracting screen; this will be shown below in deriving (16). The spatial correlogram $\rho_1(\alpha,\beta)$ of $E_1(x,y)$ is given by the usual relation

$$\rho_1(\alpha,\beta)=\frac{\overline{E_1(x,y)E_1(x+a,y+\beta)}}{\overline{E_1^2(x,y)}} \tag{13}$$

since the mean value of $E_1(x,y)$ is zero. On comparing this definition with (9), it will be seen that (13) has the term $\overline{E_1^2(x,y)}$ in the denominator, whereas, from (8), $\overline{A^2(x_0,y_0)}=1$. In fact, the mean square fluctuation in $A(x_0,y_0)$ at the screen appears, at some distance away, as a fluctuation partly in $E_2(x,y)$. It is easily shown from (12) that

$$\overline{E_1^2(x,y)}+\overline{E_2^2(x,y)}=\overline{A^2(x_0,y_0)}=1, \tag{14}$$

the sum of the mean square values of $E_1(x,y)$ and $E_2(x,y)$ remaining constant at all distances from the screen. These mean square values represent, in the case of an amplitude screen, the amount of fluctuation of the amplitude and the phase respectively at a distance z from the screen.

The autocovariance of E_1 can be calculated from the spectrum $W_1(\nu,\chi)$ in (12) by the Wiener-Khintchine theorem; $\overline{E_1^2(x,y)}$ is then its value for $\alpha=\beta=0$. Before this can be calculated, the power spectrum $W_1(\nu,\chi)$ of the signal $A(x_0,y_0)$ at the screen must be found. Some assumption is needed for the spatial characteristics of the signal. They will be assumed to be as follows:

i. $A(x,y)$ has a Gaussian correlogram in all directions.

ii. The contours of the two-dimensional correlogram are elliptical.

iii. The ellipses have their principal axes oriented along the x and y axes of the co-ordinate system.

The $\rho=0.61$ contour on the two-dimensional correlogram is used to define two structure sizes D_1 and D_2 of the screen in the x and y directions. The correlogram of $A(x,y)$ is then

$$\rho_0(\alpha,\beta)=\exp\{-(\alpha^2/2D_1^2)-(\beta^2/2D_2^2)\}. \tag{15}$$

The angular spectrum at the screen is given by an equation like (2)

$$P_0(s,r)=\iint_{-\infty}^{\infty}A(x_0,y_0)\cdot\exp\{-2\pi i(sx_0+ry_0)/\lambda\}dx_0dy_0$$

and the power spectrum is given by

$$\begin{aligned} W_0(s,r)=\overline{P_0(s,r)\cdot P_0^*(s,r)}&=\text{mean of}\left\{\iint_{-\infty}^{\infty}A(x_0,y_0)\cdot\exp\{-2\pi i(sx_0+ry_0)/\lambda\}dx_0dy_0\right.\\ &\left.\times\iint_{-\infty}^{\infty}A(x_0',y_0')\cdot\exp\{2\pi i(sx_0'+ry_0')/\lambda\}dx_0'dy_0'\right\}\\ &=\iiiint_{-\infty}^{\infty}A(x_0,y_0)\cdot A(x_0',y_0')\cdot\exp\{2\pi i[(x_0'-x_0)s+(y_0'-y_0)r]/\lambda\}dx_0dy_0dx_0'dy_0'\end{aligned}$$

where x_0' and y_0' are variables of integration, analogous to x_0 and y_0, but differentiated from them to avoid confusion. So

$$W_0(s,r)=\iiiint_{-\infty}^{\infty} A(x_0,y_0)\cdot A(x_0+\alpha,y_0+\beta).\exp\{2\pi i(\alpha s+\beta r)/\lambda\}dx_0dy_0d\alpha\, d\beta$$

where $\alpha=x_0'-x_0$, $\beta=y_0'-y_0$. By (14), it is evident that

$$\overline{A(x_0,y_0)\cdot A(x_0+\alpha,\, y_0+\beta)}=\rho_0(\alpha,\beta)$$

and therefore

$$W_0(s,r)=\iiiint_{-\infty}^{\infty}\rho_0(\alpha,\beta)\cdot\exp\{2\pi i(\alpha s+\beta r)/\lambda\}dx_0dy_0d\alpha\, d\beta.$$

Now neither x_0 nor y_0 appears in the integrand; so the integration with respect to x_0 and y_0 will give a $W_0(s,r)$ which is proportional to the total area S of the diffracting screen:

$$W_0(s,r)=S\cdot\iint_{-\infty}^{\infty}\rho_0(\alpha,\beta)\cdot\exp\{2\pi i(\alpha s+\beta r)/\lambda\}d\alpha d\beta. \tag{16}$$

Substituting for $\rho_0(\alpha,\beta)$ from (15) and integrating,

$$W_0(s,r)=2\pi SD_1D_2.\exp\{-2\pi^2(D_1^2s^2+D_2^2r^2)/\lambda\}\cdot \tag{17}$$

The spatial power spectra of $E_1(x,y)$ and $E_2(x,y)$ are found by substituting from (17) into (12):

$$\left.\begin{aligned}W_1(\nu,\chi)&=2\pi SD_1D_2.\exp\{-2\pi^2(D_1^2\nu^2+D_2^2\chi^2)/\lambda^2\}.\cos^2\{\pi z(\nu^2+\chi^2)/\lambda\}\\ W_2(\nu,\chi)&=2\pi SD_1D^2.\exp\{-2\pi^2(D_1^2\nu^2+D_2^2\chi^2)/\lambda^2\}.\sin^2\{\pi z(\nu^2+\chi^2)/\lambda\}\end{aligned}\right\} \tag{18}$$

It is now possible to find the numerator of (13); the covariance of $E_1(x,y)$ and $E_1\,(x+\alpha,y+\beta)$. The inverse transform corresponding to (16) is

$$\overline{E_1(x,y)\,E_1(x+\alpha,y+\beta)}=\frac{1}{S\lambda^2}\iint_{-\infty}^{\infty}W_1(\nu,\chi)\cdot\exp\{-2\pi i(\nu\alpha+\chi\beta)/\lambda\}d\nu\, d\chi \tag{19}$$

with the value of $W_1(\nu,\chi)$ given by (18). The integral is evaluated by putting

$$\cos^2\{\pi z(\nu^2+\chi^2)/\lambda\}=1/2+(1/4)\exp\{2\pi iz(\nu^2+\chi^2)/\lambda\}+(1/4)\exp\{-2\pi iz(\nu^2+\chi^2)/\lambda\}. \tag{20}$$

As (19) is the inverse of (16), the term 1/2 gives simply $(1/2)\rho_0(\alpha,\beta)$. The two remaining terms in (40) are identical except that the sign of z is reversed. The integral can be simplified further by separating completely the integrals with respect to ν and χ. This gives

$$\overline{E_1(x,y)\cdot E_1(x+\alpha,y+\beta)}=\frac{\pi D_1D_2}{4\lambda^2}\{I(\alpha,D_1,z)I(\beta,D_2,z)+I(\alpha,D_1,-z)I(\beta,D_2,-z)\}+(1/2)\rho_0(\alpha+\beta) \tag{21}$$

where

$$I(\alpha,D_1,z)=\int_{-\infty}^{\infty}\exp\{2\pi i[\alpha\nu+(z+\pi iD_1^2/\lambda)\nu^2]/\lambda\}d\nu$$

$$=\frac{\lambda(1+a_1^2)^{-1/4}}{\sqrt{2\pi D_1}}\exp\{-(1/2)\tan^{-1}a_1-(1-ia_1)\alpha^2/2D_1^2(1+a_1^2)\} \tag{22}$$

where the important substitutions

$$a_1=\frac{\lambda z}{\pi D_1^2},\ a_2=\frac{\lambda z}{\pi D_2^2}$$

have been made. These parameters will appear as the measures of distance on all the graphs depicting the behavior of the various statistical parameters of the diffraction pattern.

Finally, on substituting from (22) and (21),

$$\overline{E_1(x,y)\cdot E_1(x+\alpha,y+\beta)}=(1/2)\exp\{-\alpha^2/2D_1^2-\beta^2/2D_2^2\}$$

$$\underset{*}{+}(1/2)(1+a_1^2)^{-1/4}(1+a_2^2)^{-1/4}\exp\{-\alpha^2/2D_1^2(1+a_1^2)-\beta/2D_2^2(1+a_2^2)\}$$

$$\times\cos\{a_1\alpha^2/2D_1^2(1+a_1^2)+a_2\beta^2/2D_2^2(1+a_2^2(1+a_2^2)-(\tan^{-1}a_1+\tan^{-1}a_2)/2\}. \tag{23}$$

By putting $\alpha=\beta=0$

$$\overline{E_1^2(x,y)}=1/2\underset{*}{+}(1/2)(1+a_1^2)^{-1/4}\cdot(1+a_2^2)^{-1/4}\cdot\cos\{(\tan^{-1}a_1+\tan^{-1}a_2)/2\}. \tag{24}$$

The correlogram $\rho_1(\alpha,\beta)$ is given by the ratio of (23) to (24). The expressions for the autocovariance of $E_2(x,y)$ are identical, except that the positive signs, denoted above with asterisks, become negative. Some special cases will now be investigated.

(i) *Near the screen* $a_1=a_2=0$, and

$$\overline{E_1(x,y)\cdot E_1(x+\alpha,y+\beta)}=\exp(-\alpha^2/2D_1^2-\beta^2/2D_2^2)=\rho_0(\alpha,\beta)$$

$$\overline{E_1^2(x,y)}=1,\ \overline{E_2^2(x,y)}=0.$$

as would be expected, since $E_1(x,y)=A(x_0,y_0)$ at the screen.

(ii) *Very far from the screen* $a_{1,2}\to\infty$, $\tan^{-1}a_1=\tan^{-1}a_2=\pi/2$, and

$$\overline{E_1(x,y)E_1(x+\alpha,y+\beta)}=(1/2)\exp\{-\alpha^2/2D_1^2-\beta^2/2D_2^2\}$$

$$\overline{E_1'^2(x,y)}=\overline{E_2'^2(x,y)}=1/2$$

$$\rho_1(\alpha,\beta)=\rho_2(a,\beta)=\exp\{-\alpha^2/2D_1^2-\beta^2/2D_2^2\}. \tag{25}$$

The second result in (25) means that the space correlation of both amplitude and phase at a great distance from an amplitude screen is just the same as the correlation at the screen. This is *not* true at intermediate distances, as we shall shortly see. The first result in (25) shows that the fluctuations occur equally in amplitude and phase under these conditions. The phase-amplitude locus of the instantaneous signal at large distances therefore forms a circular "cloud" around the vector denoting the mean carrier signal, as in figure 1.

(iii) *Isotropic screen* $D_1=D_2=D$, and $a_1=a_2=a$.

$$\overline{E_1(x,y)E_1(x+\alpha,y+\beta)}=(1/2)\exp\{-(\alpha^2+\beta^2)/2D^2\}$$

$$+(1/2)(1+a^2)^{-1/2}\cdot\exp\{-(\alpha^2+\beta^2)/2D^2(1+a^2)\}\cdot\cos\{a(\alpha^2+\beta^2)/2D^2(1+a^2)-\tan^{-1}a\}. \tag{26}$$

The quantities α and β appear in this expression only as $(\alpha^2+\beta^2)$, so the contours of correlation are circles at all distances, and the diffraction pattern is never anisotropic. Also

$$\overline{E_1'^2(x,y)}=1/2\underset{*}{+}(1/2)(1+a^2)^{-1}$$

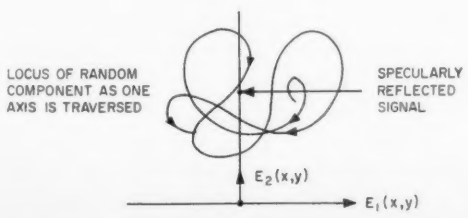


FIGURE 1. *A possible phase—amplitude diagram for the diffracted signal a long way from a random screen.*

the asterisk signifying, as previously, the sign which is to be reversed to give the corresponding value for $\overline{E_2^2(x,y)}$. $\overline{E_1^2(x,y)}$ and $\overline{E_2^2(x,y)}$ are plotted on figure 2. It can be seen that the mean square values of E_1 and E_2 are virtually equal at distances from the screen greater than that corresponding to about $a=4$. It will be shown later that, under these circumstances, the diffraction pattern is not correlated with the signal at the screen.

Slight rearrangement of (26) gives

$$\overline{E_1(x,\ y)\ E_1(x+\alpha,\ y+\beta)}=(1/2).\ \exp\{-(\alpha^2+\beta^2)/2D^2\}$$

$$\underset{*}{+}(1/2)(1+a^2)^{-1}.\ \exp\{-(\alpha^2+\beta^2)/2D^2(1+a^2)\}\ [\cos\{a(\alpha^2+\beta^2)/2D^2(1+a^2)\}$$

$$+\ a\ \sin\ \{a(\alpha^2+\beta^2)/2D^2(1+a^2)\}]. \quad (27)$$

(iv) *One-dimensional screen* $\quad D_1=D;\ D_2\to\infty$. So $a_2\to 0$.
From (24),

$$\overline{E_1^2(x,\ y)}=1/2\underset{*}{+}(1/2)(1+a^2)^{-1/4}\cdot\cos\{(\tan^{-1}\ a)/2\}$$

$$=1/2\underset{*}{+}(1/2)\{[1+(1+a^2)^{1/2}][1+a^2]^{-1}/2\}^{1/2}.$$

This also is plotted in figure 2, for comparison with the curve for the two-dimensional case. Evidently the mean square values of $E_1(x,y)$ and $E_2(x,y)$ become nearly equal much closer to a two-dimensionally irregular screen than a screen which is irregular in one dimension only. This illustrates that a one-dimensional model may give a quite inadequate representation of two-dimensional diffraction phenomena. For this case,

$$\overline{E_1(x,\ y)\ E_1(x+\alpha,\ y+\beta)}=(1/2)\ \exp(-\alpha^2/2D^2)$$

$$\underset{*}{+}(1/2)(1+a^2)^{-1'4}\ \exp\ \{-\alpha^2/2D^2(1+a^2)\}\cdot\cos\{a\alpha^2/2D^2(1+a^2)$$

$$-(\tan^{-1}\ a)/2\}.$$

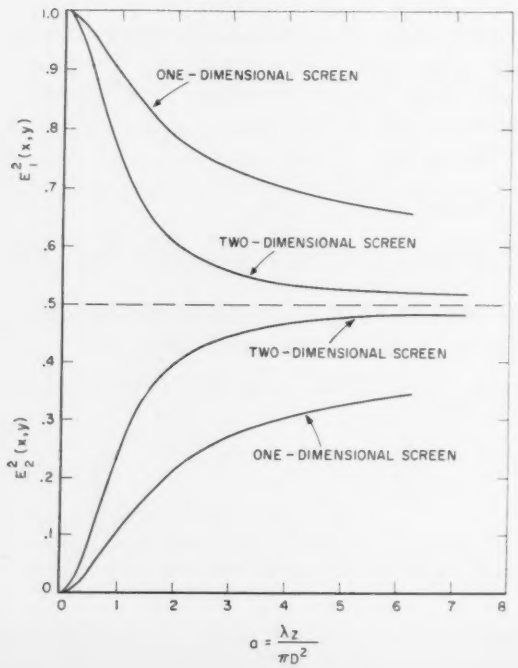


FIGURE 2. *Variation of mean square diffracted signal with distance from two kinds of random screen.*

It will be noted from (27) that the correlogram is not Gaussian in form, though $E_1(x,y)$ has a Gaussian correlogram for $a=0$, and both $E_1(x,y)$ and $E_2(x,y)$ have Gaussian correlograms as a $\rightarrow\infty$ (see equation (25)). This introduces some difficulty as to what is to be called the "structure size" d of the diffraction pattern, for comparison with experiment. The following definition is adopted

$$d=\text{Structure size in } x \text{ direction}=(-\partial^2\rho/\partial\alpha^2)^{-1/2}_{\alpha=\beta=0}$$
$$=D_1 \text{ for the correlogram } \rho_0(\alpha,\beta).$$

This definition bases the structure size on the Gaussian curve which is the best fit to the correlogram, near the origin of α. It can best be found by expanding (23) in powers of α^2 and β^2, and neglecting terms of order higher than the second; using equations (13) and (24),

$$d_1^2/D_1^2=\frac{1+(1+a_1^2)^{-1/4}(1+a_2^2)^{-1/4}\cos\{(\tan^{-1}a_1+\tan^{-1}a_2)/2\}}{1+(1+a_1^2)^{-5/4}(1+a_2^2)^{-1/4}[\cos(\tan^{-1}a_1+\tan^{-1}a_2)/2-a_1\sin\{(\tan^{-1}a_1+\tan^{-1}a_2)/2\}]}$$

$$d_2^2/D_2^2=\frac{1+(1+a_1^2)^{-1/4}(1+a_2^2)^{-1/4}\cos\{(\tan^{-1}a_1+\tan^{-1}a_2)/2\}}{1+(1+a_1^2)^{-1/4}(1+a_2^2)^{-5/4}[\cos(\tan^{-1}a_1+\tan^{-1}a_2)/2-a_2\sin\{(\tan^{-1}a_1+\tan^{-1}a_2)/2\}]}. \tag{28}$$

To help in understanding the significance of these results, a number of special cases are examined below.

(v) *Near the screen,* a_1 and $a_2\rightarrow 0$, and $d_1\rightarrow D_1$, $d_2\rightarrow D_2$ for $E_1(x,y)$, as found previously under (i). For the quadrature component $E_2(x,y)$ however, it is easily shown that

$$\left.\begin{aligned} d_1&=D_1\left\{\frac{K^4+2K^2/3+1}{K^4+2K^2+5}\right\}^{1/2}\\ d_2&=D_2\left\{\frac{K^4+2K^2/3+1}{5K^4+2K^2+1}\right\}^{1/2}\end{aligned}\right\} \tag{29}$$

where $K=D_1/D_2$.

(vi) *Isotropic screen* $D_1=D_2=D$, $K=1$.

Near the screen, $d=D$ for $E_1(x,y)$ as found previously.

From (29) $d=D/\sqrt{3}$ for $E_2(x,y)$.

This result implies that the structure size of the shallow phase variations *near* an amplitude screen (or of the amplitude variations near the phase screen) is $1/\sqrt{3}$ times the structure of the screen itself. In addition, the correlogram of $E_2(x,y)$ is far from Gaussian; it can easily be shown by expanding (23) in powers of a that, near the screen,

$$\rho_2(\alpha,\beta)=\exp\{-(\alpha^2+\beta^2)/2D^2\}\cdot\{1-(\alpha^2+\beta^2)/D^2+(\alpha^2+\beta^2)^2/8D^4\} \tag{30}$$

exactly. This is plotted in figure 3, and has an oscillating form. At any distance from an isotropic screen, (28) reduces to for $E_1(x,y)$:

$$d=D\left\{\frac{(1+a^2)(2+a^2)}{a^4+a^2+2}\right\}^{1/2}$$

for $E_2(x,y)$:

$$d=D\left\{\frac{1+a^2}{3+a^2}\right\}^{1/2}.$$

These relations are shown plotted in figure 4. The structure size of $E_1(x,y)$ actually has a maximum at intermediate values of z.

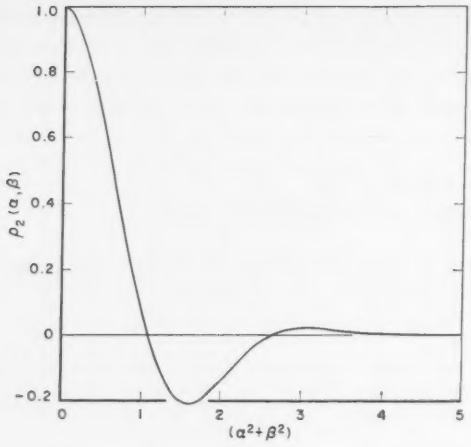


FIGURE 3. *Correlogram of the shallow phase variations near a two-dimensional amplitude screen.*

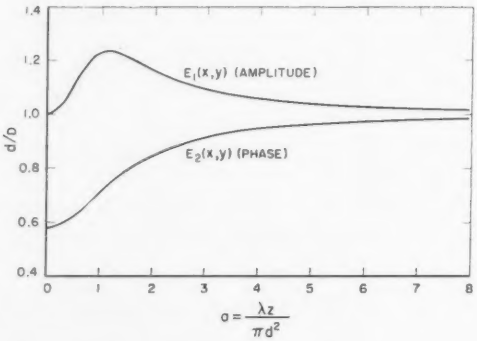


FIGURE 4. *Variation of structure size* d *with distance from a two-dimensional amplitude screen.*

(vii) *One-dimensional screen* $D_1=D$; $D_2\to\infty$, $a_2\to 0$, $K\to 0$. Near the screen,

$$d=D \text{ for } E_1(x,y),\quad d=D/\sqrt{5} \text{ for } E_2(x,y),$$

showing the effect described in (vii) to an even greater extent. At any distance, for $E_1(x,y)$:

$$d=D\left\{\frac{1+(1+a^2)^{-1/4}\cos\{(\tan^{-1}a)/2\}}{1+(1+a^2)^{-5/4}\,[\cos\{(\tan^{-1}a)/2\}-a\sin\{(\tan^{-1}a)/2\}]}\right\}^{1/2}$$

for $E_2(x,y)$:

$$d=D\left\{\frac{1-(1+a^2)^{-1/4}\cos\{(\tan^{-1}a)/2\}}{1-(1+a^2)^{-5/4}\,[\cos\{(\tan^{-1}a)/2\}-a\sin\{(\tan^{-1}a)/2\}]}\right\}^{1/2}.$$

In the general case of an anisotropic screen, (28) can be used to find the structure sizes d_1 and d_2 in the x and y directions. To clarify the kind of behavior that may occur, figure 5 has been prepared. This shows how the contours of the correlation ellipse change shape with increasing distance from a random amplitude screen. The two screens shown are

(a) isotropic, with $D_1=D_2=1$

(b) anisotropic, with $D_1=2$, $D_2=1$.

The distances from the screen were chosen as 0, ∞, and $\pi D_1^2/\lambda$. It is interesting to note that the phase correlation ellipse for the anisotropic screen does not keep the same shape as the screen is approached—its minor axis is reduced by a greater factor than its major axis. This is because as it approaches the limiting case of an anisotropic screen—namely, a one-dimensional screen—the major axis of the phase correlation ellipse near the screen approaches the true structure size D_1 of the screen, while the minor axis is reduced by the factor $\sqrt{5}$ established above in (vii).

The reason for dealing with this topic in such detail is to demonstrate that the structure size of the irregularities in the diffraction pattern produced by a shallow phase or amplitude

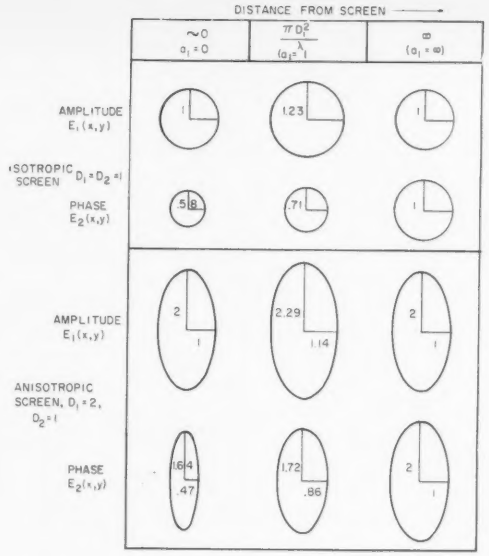


FIGURE 5. *Contours of the spatial correlogram of the wave diffracted from two kinds of amplitude screen, at different distances.*

modulated screen is not always the same as that of the screen itself. The preceding lengthy mathematical argument is necessary to establish the precise effect of the diffraction process, so that existing experimental data may be interpreted with confidence.

4. Cross-Correlation Relationships for E_1 and E_2

A good deal of attention has been paid by various workers (for instance, Jones, Millman, and Nertney [1953]) to the question of whether the signal in the diffraction pattern is the same as the signal at a random screen directly opposite to it. It must be emphasized that the question is not whether the signals are *statistically* similar, i.e., have the same structure size or fading time, but whether the *detailed* changes in signal are similar at the two planes. Expressed in more exact language, the cross-correlation is to be evaluated between, for instance, the amplitude of the signal at the screen and the amplitude at a distant plane.

Another cross-correlation coefficient of considerable importance is that between the amplitude and the phase of the diffracted signal [Bowhill, 1957]. Jones, Millman, and Nertney [1953] assumed that this correlation arises as a result only of correlation between the amplitude and phase of the signal at the screen. It is shown here, however, that they may be correlated even for a screen which changes the phase only or the amplitude only.

The theory is developed for a shallow amplitude modulating screen, and is extended to the case of a phase screen. The quantity $E_1(x,y)$ therefore represents the amplitude of the signal in the diffraction pattern (with the specular component removed) and $E_2(x,y)$ represents the phase of the signal. The quantities

$$\overline{A(x,y)E_1(x,y)} \text{ and } \overline{E_2(x,y)E_1(x,y)} \tag{31}$$

are to be calculated; the known values of $\overline{E_1^2(x,y)}$, $\overline{E_2^2(x,y)}$ and $\overline{A^2(x,y)}$ then give the cross-correlation coefficients ρ_{11} between $A(x,y)$ and $E_1(x,y)$, and ρ_{12} between $E_1(x,y)$ and $E_2(x,y)$.

The quantities (31) are computed from the relations

$$\left.\begin{aligned}\overline{A(x,y)E_1(x,y)} &= (1/2)\overline{E_1^2(x,y)} + (1/2)\overline{A^2(x,y)} - (1/2)\overline{[A(x,y)-E_1(x,y)]^2}\\ \overline{E_1(x,y)E_2(x,y)} &= (1/2)\overline{E_1^2(x,y)} + (1/2)\overline{A^2(x,y)} - (1/2)\overline{[E_1(x,y)-E_2(x,y)]^2}.\end{aligned}\right\} \tag{32}$$

The last terms in these equations represent the areas under the power spectra of $[A(x,y)-E_1(x,y)]$ and $[E_1(x,y)-E_2(x,y)]$, since the mean square value of any random variable with zero mean is given by the integral of its power spectrum with respect to frequency.

The power spectrum of $[A(x,y)-E_1(x,y)]$ can be found from (11);

$$P_1(\nu,\chi)=P_0(\nu,\chi)\cdot\cos\{\pi z(\nu^2+\chi^2)/\lambda\}. \tag{11}$$

The spectrum of $[A(x,y)-E_1(x,y)]$ is therefore

$$P_0(\nu,\chi)-P_1(\nu,\chi)=P_0(\nu,\chi)[1-\cos\{\pi z(\nu^2+\chi^2)/\lambda\}]$$

and its power spectrum is given, exactly as in (6), by

$$W_0(\nu,\chi)[1-\cos\{\pi z(\nu^2+\chi^2)/\lambda\}]^2=4\,[W_2(\nu,\chi)]_{z/2}-[W_2(\nu,\chi)]_z$$

and substituting in (32),

$$\overline{A(x,y)E_1(x,y)}=(1/2)[\overline{E_1^2(x,y)}]_z+(1/2)-2\,[\overline{E_2^2(x,y)}]_{z/2}+(1/2)[\overline{E_2^2(x,y)}]_z=2\,[\overline{E_1^2(x,y)}]_{z/2}-1$$

a relation which holds for any autocorrelation function at the screen. Using $\overline{E_1^2(x,y)}$ from (24), the correlation ρ_{11} is given by

$$\rho_{11}=\overline{A(x,y)E_1(x,y)}/\overline{E_1^2(x,y)}=\frac{\sqrt{2}\cdot(1+a_1^2/4)^{-1/4}(1+a_2^2/4)^{-1/4}\cos\{[\tan^{-1}(a_1/2)+\tan^{-1}(a_2/2)]/2\}}{\{1+(1+a_1^2)^{-1/4}(1+a_2^2)^{-1/4}\cos\{[(\tan^{-1}a_1+\tan^{-1}a_2)/2]\}\}^{1/2}}$$

for a screen with a Gaussian correlogram.

For an isotropic screen, $a_1=a_2=a$, and this reduces to

$$\rho_{11}=\{(1+a^2)/(1+a^2/2)\}^{1/2}(1+a^2/4)^{-1}.$$

This function is plotted as the full curve in figure 6. The correlation is perfect near the screen, since then $E_1(x,y)=A(x,y)$. It decreases progressively to zero with increasing distance, reaching a value of 0.1 when $a=7.2$.

The corresponding function for a one-dimensional screen, $a_1=a$, $a_2=0$, is given by

$$\rho_{11}=\frac{2(1+a^2/4)^{-1/4}\cos\{[\tan^{-1}(a/2)]/2\}}{\{1+(1+a^2)^{-1/4}\cos\{(\tan^{-1}a_1+\tan^{-1}a_2)\}\}^{1/2}}$$

This function is plotted as the broken curve on figure 6. The correlation decreases much less rapidly with increasing distance from the screen than for the two-dimensional case.

To find the correlation ρ_{12}, (32) is rearranged to give

$$\overline{E_1(x,y)E_2(x,y)}=1/2-(1/2)[\overline{E_1(x,y)-E_2(x,y)}]^2$$

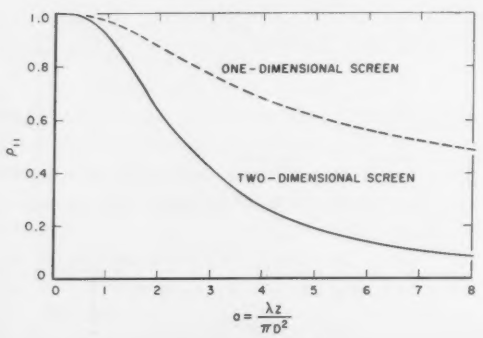


FIGURE 6. *Variation of the correlation of the diffraction pattern amplitude with the amplitude at the screen as a function of distance for one- and two-dimensional screens.*

and the power spectrum of $[E_1(x,y)-E_2(x,y)]$ is given from (11) by

$$\overline{[P_1(\nu,\chi)-P_2(\nu,\chi)][P_1^*(\nu,\chi)-P_2^*(\nu,\chi)]}$$

$$=W_0(\nu,\chi)[\cos\{\pi z(\nu^2+\chi^2)/\lambda\}-\sin\{\pi z(\nu^2+\chi^2)/\lambda\}]^2=W_0(\nu,\chi)-W_0(\nu,\chi)\sin\{2\pi z(\nu^2+\chi^2)/\lambda\}.$$

So

$$\overline{E_1(x,y)E_2(x,y)}=(1/2\lambda^2S)\iint_{-\infty}^{\infty}W_0(\nu,\chi)\sin\{2\pi z(\nu^2+\chi^2)/\lambda\}d\nu d\chi$$

$$=(\pi D_1D_2/2i\lambda^2)\iint_{-\infty}^{\infty}[\exp\{2\pi iz(\nu^2+\chi^2)/\lambda\}-\exp\{-2\pi iz(\nu^2+\chi^2)/\lambda\}]$$

$$\cdot\exp\{-2\pi^2(D_1^2\nu^2+D_2^2\chi^2)/\lambda^2\}d\nu d\chi.$$

$$=(\pi D_1D_2/2i\lambda^2)\left\{\int_{-\infty}^{\infty}\exp\{2\pi i\nu^2(z+\pi iD_1^2/\lambda)/\lambda\}d\nu\int_{-\infty}^{\infty}\exp\{2\pi i\chi^2(z+\pi iD_2^2/\lambda)/\lambda\}d\chi\right.$$

$$\left.-\int_{-\infty}^{\infty}\exp\{2\pi i\nu^2(-z+\pi iD_1^2/\lambda)/\lambda\}d\nu\int_{-\infty}^{\infty}\exp\{2\pi i\chi^2(-z+\pi iD_2^2/\lambda)/\lambda\}d\chi\right\}$$

and on integrating this expression by the usual methods

$$\overline{E_1(x,y)E_2(x,y)}=(1/2)(1+a_1^2)^{-1/4}(1+a_2^2)^{-1/4}\sin\{(\tan^{-1}a_1+\tan^{-1}a_2)/2\}.$$

Also, from (24)

$$4\cdot\overline{E_1^2(x,y)}\cdot\overline{E_2^2(x,y)}=1-(1+a_1^2)^{-1/2}(1+a_2^2)^{-1/2}\cos^2\{(\tan^{-1}a_1+\tan^{-1}a_2)/2\}$$

and the correlation ρ_{12} between $E_1(x,y)$ and $E_2(x,y)$ is given by

$$\rho_{12}=\frac{(1+a_2^2)^{-1/4}(1+a_2^2)^{-1/4}\sin\{(\tan^{-1}a_1+\tan^{-2}a_2)/2\}}{[1-(1+a_1^2)^{-1/2}(1+a_2^2)^{-1/2}\cos^2\{(\tan^{-1}a_1+\tan^{-1}a_2)/2\}]^{1/2}}. \tag{33}$$

For an isotropic screen, $a_1=a_2=a$; this reduces to

$$\rho_{12}=(1+a^2)^{-1/2}$$

which is plotted in figure 7. The correlation between $E_1(x,y)$ and $E_2(x,y)$ has a maximum value of $1/\sqrt{2}$ at the screen, and decreases to zero with increasing distance. For an anisotropic screen, (33) must be expanded in powers of a_1 and a_2. The correlation as a_1 and $a_2\to 0$ is given by

$$\rho_{12}=\frac{a_1+a_2}{[3a_1^2+2a_1a_2+3a_2^2]^{1/2}}=\frac{1+K^2}{[3K^4+2K^2+3]^{1/2}} \tag{34}$$

where $K=D_1/D_2$ expresses the degree of anisotropy of the screen, as in (29). Whatever the value of K, (34) shows that this maximum value of the correlation ρ_{12} lies between the values

$$\rho_{12}=1/\sqrt{2} \text{ for an isotropic screen } a_1=a_2$$

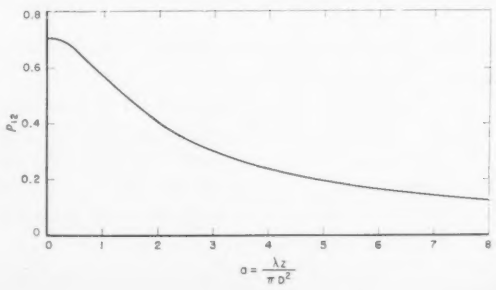


FIGURE 7. *Variation of amplitude-phase cross-correlation with distance from a two-dimensional amplitude screen.*

and

$$\rho_{12}=1/\sqrt{3} \text{ for a one-dimensional screen } a_1 \gg a_2. \tag{35}$$

The cross-correlation coefficients ρ_{11} and ρ_{12} between the phase components $E_1(x,y)$ and $E_2(x,y)$ must now be related to the amplitude and phase characteristics of the diffraction pattern. For an amplitude screen, it is apparent from figure 1 that for a shallow screen,

$$\text{Amplitude in pattern} \triangleq (\text{specular signal}) + E_1(x,y)$$
$$\text{Phase in pattern} \triangleq -E_2(x,y)/(\text{specular signal})$$

since the positive phase direction is opposite to the positive E_2 direction, taking an increase in phase to mean a *retarding* of the wave. This means that:

Correlation between amplitude in pattern and at screen $=\rho_{11}$, and correlation between amplitude and phase in pattern $=-\rho_{12}$. Similarly, for a *phase modulating screen*

$$\text{Amplitude in pattern} \triangleq (\text{specular signal}) + E_2(x,y)$$
$$\text{Phase in pattern} \triangleq E_1(x,y)/(\text{specular signal})$$

and also

$$\text{Correlation between phase in pattern and at screen} = \rho_{11}$$
$$\text{Correlation between phase and amplitude in pattern} = \rho_{12}.$$

It has therefore been established that the correlation between the amplitude and phase in a diffracted signal is zero at large distances, but at smaller distances is positive for a phase modulating screen and negative for an amplitude screen. This occurs at a distance from the screen given by $a \sim 1$, or $a \sim D^2/\lambda$. This is closely related to the Rayleigh distance for an optical system of aperture equal to the structure size of the screen irregularities.

This result has been derived by setting up the wave solution. It may be partially confirmed by a physical-optical argument, as follows. Consider a one-dimensional phase-modulating screen, with a phase profile $\phi(x)$. The signal at a point near the screen is determined only by the properties of the screen in its immediate neighborhood. In fact, any small portion of the screen centered about x_0 acts as a positive lens of focal length

$$F=-2\pi/\lambda \left[\frac{d^2\phi}{dx^2}\right]_{x=x_0}. \tag{36}$$

The plane wave incident on the screen can be imagined as coming from a point source of waves, very far away. The image formed by the effective "lens" of the phase screen appears to have angular magnification $=(1-z/F)^{-1}$ where z is the distance of the point of observation from the screen. The image of the source appears the same brightness as the source itself, so the signal power arriving at z is proportional to the angular magnification. So the signal amplitude is proportional to $(1-z/F)^{-1/2}$; if z is assumed much less than F, approximately

$$\text{signal amplitude} \propto 1+z/2F=1-(\lambda z/4\pi)\left[\frac{d^2\phi}{dx^2}\right]_{x=x_0}$$

on substitution from (36). Any correlation between the amplitude and phase in the pattern must therefore arise through correlation between $\phi(z)$ and $(-d^2\phi/dx^2)$.

In fact, it can easily be shown that if $\phi(x)$ is a Gaussianly distributed random function, there is a correlation between these two quantities of $[\rho''(0)/\rho''''(0)]^{1/2}$, where $\rho(\alpha)$ is the correlogram of $\phi(x)$. For a Gaussian correlogram, which has been assumed throughout this section, this reduces to $1/\sqrt{3}$; a result which is identical with that given for the same case in (35). It should be emphasized, however, that geometric optics can *only* be used near the screen, where the diffraction effects have not yet developed appreciably.

5. Effect of a Point Source of Waves

In all the previous sections it has been assumed that there is no systematic variation of phase across the random screen. However, if the screen is illuminated with radiation from a point source a distance z' from the screen, instead of from a plane wave source, the phase of the wave before passing through the screen varies with distance. The signal emerging from the screen is now given by

$$E_0(x_0,y_0)=A(x_0,y_0)\ \exp\ \{2\pi i(z'^2+x_0^2+y_0^2)^{1/2}/\lambda\}.$$

If the cone angle of the diffracted waves is small, the important regions of the screen are for $x_0,\ y_0 \ll z'$. So the approximation

$$(z'^2+x_0^2+y_0^2)^{1/2} \triangleq z'+x_0^2/2z'+y_0^2/2z' \tag{37}$$

may be used.

Exactly as in section 3, the vector signal $\underline{E}(x,y,z)$ at a distance z from the screen can be found and the phase components $E_1(x,y,z)$ and $E_2(x,y,z)$ evaluated by a relationship similar to (9). In that equation, the phase of the specular component is independent of x and y since the specular component is a plane wave. In the present case, however, the specular wave is spherical and so the analogous equation to (9) is

$$\underline{E}(x,y,z)=[E_1(x,y,z)+iE_2(x,y,z)]\ \exp\ \{2\pi i[(z+z')^2+x^2+y^2]^{1/2}/\lambda\}$$

and for $x,y \ll z$,

$$[(z+z')^2+x^2+y^2]^{1/2} \triangleq z+z'+x^2/2(z+z')+y^2/2(z+z'). \tag{38}$$

The corresponding equation to (10) can then be written down as

$$\underline{E}(x,y,z)\ \exp\ \{-2\pi i\ [(z+z')^2+x^2+y^2]^{1/2}/\lambda\}=(1/\lambda^2)\iiiint_{-\infty}^{\infty} A(x_0,y_0)\cdot \exp\{2\pi i\ [(x-x_0)s + (y-y_0)r+(z'^2+x_0^2+y_0^2)^{1/2}+z(1-s^2/2-r^2/2)]/\lambda\}\ dx_0dy_0dsdr$$

or

$$\underline{E}(x,y,z,)=(1/\lambda^2)\iiiint_{-\infty}^{\infty} A(x_0,y_0)\ \exp\ \{2\pi iG/\lambda\}\ dx_0dy_0dsdr$$

where

$$G=(x-x_0)s+(y-y_0)r+(z'^2+z_0^2+y_0^2)^{1/2}+z(1-s^2/2-r^2/2)-[(z+z')^2+x^2+y^2]^{1/2}.$$

By substitution from (37) and (38), approximately

$$G=(x-x_0)s+(y-y_0)r+z'+x_0^2/2z'+y_0^2/2z'+z-zs^2/2-zr^2/2-z-z'-x^2/2(z+z')-y^2/2(z+z')$$
$$=(x-x_0)s+(y-y_0)r+x_0^2/2z'+y_0^2/2z'-zr^2/2-zs^2/2-x^2/2(z+z')-y^2/2(z+z').$$

The phase components $E_1(x,y,z)$ and $E_2(x,y,z)$ are given by

$$\left.\begin{aligned} E_1(x,y,z)&=\mathscr{R}\{\underline{E}(x,y,z)\}=(1/2\lambda^2)\iiiint_{-\infty}^{\infty} A(x_0,y_0)\{\exp\ (2iG/\lambda)+\exp\ (-2iG/\lambda)\}\ dx_0dy_0dsdr \\ E_2(x,y,z)&=\mathscr{I}\{\underline{E}(x,y,z)\}=(1/2\lambda^2)\iiiint_{-\infty}^{\infty} A(x_0,y_0)\{\exp\ (2iG/\lambda)-\exp\ (-2iG/\lambda)\}dx_0dy_0dsdr \end{aligned}\right\} \tag{39}$$

and the spatial spectra of $E_1(x,y,z)$ and $E_2(x,y,z)$ are given, as before, by

$$\left.\begin{aligned} P_1(\nu,\chi)&=\iint_{-\infty}^{\infty} E_1(x,y,z)\cdot\exp\{-2\pi i(\nu x+\chi y)/\lambda\}dxdy \\ P_2(\nu,\chi)&=\iint_{-\infty}^{\infty} E_2(x,y,z)\cdot\exp\{-2\pi i(\nu x+\chi y)/\lambda\}dxdy. \end{aligned}\right\} \tag{40}$$

On substituting from (39) into (40) and integrating with respect to s, r, x, and y, an analogous expression to (30) is obtained:

$$P_1(\nu,\chi)=\frac{z+z'}{z'}\cos\{\pi(\nu^2+\chi^2)z(z+z')/\lambda z'\}\cdot\iint_{-\infty}^{\infty}A(x_0,y_0)\exp\{-2\pi i(\nu x_0+\chi y_0)(z+z')/\lambda z'\}\,dx_0dy_0. \tag{41}$$

The differences between (41) and (11) can be accounted for by supposing that there are apparent values of z and λ which, when substituted in (11) *and in all the results derived from it*, just allow for the fact that there is a point source of waves at a distance z' from the screen.

In comparing (41) and (11), the double integral will first be considered. Both ν and χ in the exponent are multiplied by the factor $(z+z')/z$; so the angular spectrum must be scaled down in width by the reciprocal of this factor, relative to its width in the plane wave case.

The "generalized autocorrelation function" of Booker, Ratcliffe, and Shinn (1950) given by the Fourier transform of the angular power spectrum, is invariant with distance in the plane wave case. With point source illumination, its width increases proportionally to $(1+z/z')$, and finally increases proportionally to z. In the case of vertical incidence of waves on the ionosphere $z=z'$, and the apparent structure size of the screen irregularities is just twice the actual structure size.

These results can be appreciated physically by considering the behavior of the diffraction pattern as $\lambda\to 0$, and an analogy can be drawn with geometrical optics. If an amplitude screen (see fig. 8) is illuminated with a point source of light at a distance z' from it, the scale of the amplitude variations a distance z away from the screen will just be $(1+z/z')$ times the scale of the amplitude variations at the screen.

The other respect in which (41) and (11) differ is in the argument of the cosine. The quantity $(\nu^2+\chi^2)z(z+z')/z'$ replaces $(\nu^2+\chi^2)z$. But the quantities ν and χ have already been found to be multiplied by $(1+z/z')$; so z in (30) must be replaced by

$$\frac{z(z+z')}{z'}\,\frac{1}{(1+z/z')^2}=\frac{zz'}{z+z'}=z_{eff},$$

where z_{eff} is the effective value of z. In the case of vertical incidence on the ionosphere, $z=z'$, and $z_{eff}=z/2$.

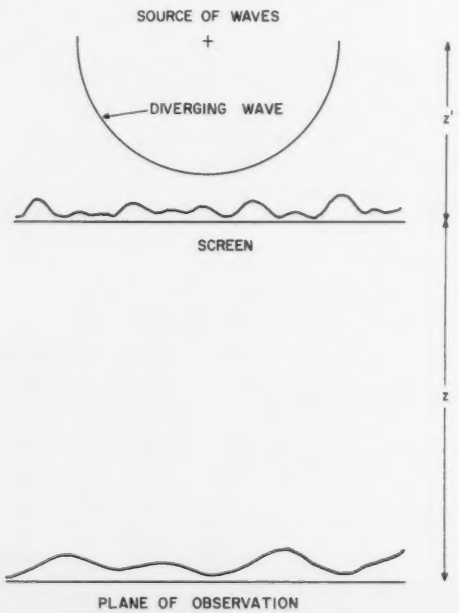


FIGURE 8. *Diagram showing the increase in structure size of the amplitude variations with increasing distance from a point source of waves.*

It is interesting to note that

$$\frac{1}{z_{eff}}-\frac{1}{z'}=\frac{1}{z}$$

which is exactly comparable with the object and image distance relations for a zone plate, the theory of which is closely related to that given above.

6. Effect of Oblique Incidence

If a plane wave is incident at an angle i to a diffracting screen, the equations corresponding to those of section 3 can be deduced by an exactly similar method. If y axis is in the plane of the incident wave, we may write $s=\sin i+s_1$, where $(\sin^{-1}s-\sin^{-1}s_1)$ is the scattering angle. The phase of the specular signal at a distance z from the screen is now

$$(2\pi iz\cos i)/\lambda$$

and (9) must therefore be written

$$\underline{E}(x,y,z)=[E_1(x,y)+iE_2(x,y)]\exp[(2\pi iz\cos i)/\lambda]. \tag{42}$$

The phase of the scattered wave at distance z is again given by

$$2\pi iz\{1+(s^{-2}-1)^{-1}+(r^2-1)^{-1}\}^{-1/2}/\lambda.$$

Putting $s=\sin i+s_1$ and expanding in powers of s_1 and r, this becomes

$$2\pi iz(\cos i-s_1\tan i-\tfrac{1}{2}s_1^2\sec^3 i-\tfrac{1}{2}r^2\cos^3 i)/\lambda$$

and the exponent in the resulting equation for (E_1+jE_2), analogous to (10), is

$$2\pi i\left\{(x-x_0-z\tan i)s_1+(y-y_0)r-\frac{1}{2}s_1^2\,z\sec^3 i-\frac{1}{2}r^2z\cos^3 i\right\}/\lambda.$$

Comparing this with the exponent in (10), it is evident that

(i) the origin of x has been effectively displaced a distance $z\tan i$ in the plane of propagation—namely, to the "mid-point of the path", in propagation terminology.

(ii) the effective distance of the observing point from the screen has been multiplied by factors of $\sec^3 i$ and $\cos^3 i$ respectively for the x and y variations in the screen.

To make this more easily comprehensible, consider the case of oblique incidence on a diffracting screen with an isotropic spatial correlation function, Gaussian in form. The structure size of the interference pattern is found by putting

$$\left.\begin{aligned}a_1&=a\sec^3 i\\ a_2&=a\cos^3 i\end{aligned}\right\}\qquad K=\sec^6 i$$

in (28). Near the screen, from (29),

$$d_1=D\left\{\frac{\sec^{24}i+(2/3)\sec^{12}i+1}{\sec^{24}i+2\sec^{12}i+5}\right\}^{1/2}$$

$$d_2=D\left\{\frac{\sec^{24}i+(2/3)\sec^{12}i+1}{5\sec^{24}i+2\sec^{12}i+1}\right\}^{1/2}.$$

Another important effect of oblique incidence, which does not enter explicitly into this analysis, is the following. If the random screen is composed of a "buckled specular reflector", it will act as a phase screen for normally incident waves. However, at oblique incidence, the phase change it causes in the wave is scaled by a factor $\cos i$. This is related to the familiar phenomenon that an optical surface becomes more nearly a specular reflector at grazing incidence. The same effect is observed for radio waves.

580999—61——7

7. Conclusion

Detailed calculations have been given for the statistics of a radio wave diffracted by a random screen. The results, some of which have been used in a previous paper [Bowhill, 1957], now take account of anisotropy of the diffracting screen, and oblique incidence of the waves

8. References

Booker, H. G., J. A. Ratcliffe, and D. H. Shinn, Diffraction from an irregular screen with application to ionospheric problems, Phil. Trans. Roy. Soc. **A 242,** 579 (1950).

Bowhill, S. A., Ionospheric irregularities causing random fading of very low frequencies, J. Atmospheric and Terrest. Phys. **11,** 91 (1957).

Chernov, L. A., Wave propagation in a random medium (McGraw-Hill Book Co., Inc. New York, 1960).

Feinstein, J., Some stochastic problems in wave propagation, Part I and Part II, Trans. I.R.E. **AP–2,** 23 and 63 (1954).

Hewish, A., The diffraction of radio waves in passing through a phase-changing ionosphere, Proc. Roy. Soc. **A 209,** 81 (1951).

Hewish, A., The diffraction of galactic radio waves as a method of investigating the irregular structure of the ionosphere, Proc. Roy. Soc. **A 214,** 494 (1952).

Jones, R. E., G. H. Millman, and R. J. Nertney, The heights of ionospheric winds as measured at long radio wavelengths, J. Atmospheric and Terrest. Phys. **3,** 79 (1953).

Pitteway, M. L. V., The reflection of radio waves from a stratified ionosphere modified by weak irregularities, Proc. Roy. Soc. **A 246,** 556 (1958).

Ratcliffe, J. A., Some aspects of diffraction theory and their application to the ionosphere, Rep. Prog. Phys. **19,** 188 (1956).

(Paper 65D3–130)

JOURNAL OF RESEARCH of the National Bureau of Standards—D. Radio Propagation
Vol. 65D, No. 3, May–June 1961

Space Analysis of Radio Signals [1, 2]

John B. Smyth

(August 22, 1960)

The radio antenna is viewed as a space frequency filter with an output just equal to the convolution of its transfer function with the radio field illuminating its aperture. An equivalent uncertainty principle limits the accuracy with which the spatial distribution of the radio field may be determined. The radio field generated by an antenna is distorted in passing into the ionosphere, generating new space frequencies which is the information contained in the field at the receiving antenna. The energy diffracted into the different orders will appear to arrive from different directions, and the angle of arrival for a given order will be a function of the radiofrequency.

1. Introduction

Radio fields are functions of space and time of the form

$$f(\omega t+\bar{k}\cdot\bar{r}),$$

where ω is 2π times the time frequency in cycles per unit time and

$$\bar{k}=\bar{i}k_x+\bar{j}k_y+\bar{k}k_z \tag{1}$$

is 2π times the space frequency [3] in cycles per unit distance along the given direction. A monochromatic plane wave propagating in a homogeneous isotropic medium will have space frequencies ranging from zero along directions in the plane of the wave front to a maximum equal to 2π divided by the wave length along the direction of propagation. For directions making an angle θ with the direction of propagation, the space frequency, or wave number becomes

$$k(\theta)=\frac{2\pi}{\lambda}\cos\theta. \tag{2}$$

If the time frequency of the plane wave is not constant, then the space frequency will not be constant along directions other than $\theta=90°$, which corresponds to zero space frequency. Thus, a radio wave modulated in the time domain has a corresponding modulation in the space domain.

Even with a source monochromatic in the time domain, there is a spectrum of space frequencies which is generated by the radiating system, and this spectrum is a unique property of the current distribution associated with the radiating structure. In a homogeneous isotropic medium, the field everywhere is described in terms of this spectrum. In nonhomogeneous media this spectrum is continually modified as the wave propagates along, the inhomogeneities generate new space frequencies which feed on the energy of the original field. Other treatments of the effects of inhomogeneities, with special application to the ionosphere, have been reviewed by Ratcliffe [1956].

2. Space Signals Developed Over Ionospheric Links (1.5–15 Mc/s)

Tropospheric effects at these frequencies will be macroscopic in nature and the gross decrease in refractive index with elevation will produce a downward bending which will amount to some 22 min at an elevation angle of 2°, and 3 min at 20°. The ionosphere on the other hand, will appear dense and turbulent to the radiation propagating in the forward direction.

If, at each turning point of the radio ray, a reflecting plane parallel to the earth is assumed, and the intervening medium is considered homogeneous, then the radio field developed at the earth is just that associated with the image of the transmitter. Since the transmitter is at a large distance from the receiver, the waves arriving are almost plane and the phase wave length in a plane tangent to the earth's surface is the ratio of the radio wave length to the cosine of the elevation angle.

The medium between the reflecting plane and the earth is nonhomogeneous. In addition to the more or less regular change of electron density with elevation there are irregular patches of ionization and, in particular, filaments left in the wake of meteors at elevations between 80 and 120 km. In propagating through these irregularities, the radio waves are distorted and the image of the transmitting antenna in the ionospheric mirror becomes blurred. There are ionospheric conditions where the ray path may indicate directions up to 180° off the great circle path between the transmitter and the receiver. In any case, all that is available to the observer is the output voltage from the antenna which is generated by the radio field in its aperture. Certain deductions are made based upon the sampling of the radio field with the antenna which is nothing more than a bandpass space frequency filter. If the antenna is rotated, this is equivalent to shifting the pass-band across the space frequency domain in a given direc-

[1] Contribution from Smyth Research Associates, 3555 Aero Court, San Diego 11, Calif.
[2] Paper presented at Conference on Transmission Problems Related to High-Frequency Direction Finding, at UCLA, June 21–24, 1960.
[3] The space frequencies are frequently called wave numbers and sometimes modes, but it is felt that the term space frequency is more appropriate in this discussion.

tion. If the radio field consists of a single plane wave, the output voltage from the antenna will be its transfer function for this specific cut in the space frequency domain. For an arbitrary radio field illumination, the output from an antenna is the convolution of the radio field with its transfer function. As the antenna rotates, there may be several peaks in the output, each corresponding to increased signal derived from certain of the space frequencies making up the space signal. These results are usually interpreted as multipath transmission, where the signal is assumed to be coming from several directions.

2.1. Radio Field Distortion in Nonhomogeneous Medium

Returning to the assumed case of a plane reflector at the ray turning point in the ionosphere, the radio field at the receiver can be considered as produced by transmission from the image transmitter through the inhomogeneous stratum, the ionosphere. Let it be assumed that the ionosphere is plane and horizontally stratified and the transmission problem is depicted in figure 1. If the separation of T' the transmitter image and R, the receiver (along the z direction) is large compared with the layer thickness $2l$, the radiation incident on the layer may be assumed plane. In this case, the solution of the field as a function of z is all that is required.

A plane wave polarized normal to the incident plane will produce a field in the inhomogeneous medium of the form

$$E_x(z,t)=A(z,t)\sin[\omega t-k_z z+\varphi(z,t)] \tag{3}$$

This field satisfies the inhomogeneous wave equation

$$\frac{d^2E_x}{dz^2}+k_z^2E_x=f\left(E_x,\frac{\partial E_x}{\partial t}\right) \tag{4}$$

If it is assumed that A and φ are slowly varying functions of z, so that to a first approximation $f\left(E_x,\frac{\partial E_x}{\partial t}\right)$ may be neglected, then the unperturbed solution becomes

$$E_x=A\sin(\omega t-k_z z+\varphi), \tag{5}$$

and

$$\frac{\partial E_x}{\partial z}=-Ak_z\cos(\omega t-k_z z+\varphi). \tag{6}$$

Differentiating (3) with respect to z and substituting for $\frac{\partial E_x}{\partial z}$ from (6) yields

$$A'\sin[\omega t-kz+\varphi]+A\varphi'\cos[\omega t-kz+\varphi]=0. \tag{7}$$

Taking the derivative of (6) with respect to z and substituting into (4) yields

$$\frac{\partial A}{\partial z}k_z\cos[\omega t-kz+\varphi]-Ak\frac{\partial\varphi}{\partial z}\sin[\omega t-kz+\varphi]=f\left(E_x,\frac{\partial E_x}{\partial t}\right). \tag{8}$$

Multiplying (7) by $\sin[\omega t-kz+\varphi]$ and (8) by $\left(\frac{1}{k}\right)\cos[\omega t-kz+\varphi]$ and adding gives

$$\frac{\partial A}{\partial z}=\frac{1}{k}f\left(E_x,\frac{\partial Ex}{\partial t}\right)\cos[\omega t-kz+\varphi]. \tag{9}$$

In a similar manner

$$\frac{\partial\varphi}{\partial z}=\frac{-1}{kA}f\left(E_x,\frac{\partial E_x}{\partial t}\right)\sin[\omega t-kz+\varphi]. \tag{10}$$

It is assumed that the rates of change of the amplitude and phase vary little in a space period $2\pi/k$. The mean values fo $\frac{\partial A}{\partial z}$ and $\frac{\partial\varphi}{\partial z}$ in one period are given by:

$$\overline{\frac{\partial A}{\partial z}}=\frac{1}{2\pi k}\int_0^{2\pi}f\left(E_x,\frac{\partial E_x}{\partial t}\right)\cos kz\,d(kz), \tag{11}$$

$$\overline{\frac{\partial\varphi}{\partial z}}=-\frac{1}{2\pi kA}\int_0^{2\pi}f\left(E_x,\frac{\partial E_x}{\partial t}\right)\sin kz\,d(kz). \tag{12}$$

If $f\left(E_x,\frac{\partial E_x}{\partial t}\right)$ is a function of E_x only, the average change in the amplitude of the radio field over a wave length is zero. On the other hand, equation (12) does not vanish, and the phase change per wave length is the first order effect associated with the penetration of the field into the inhomogeneous medium. For the case where the dominant term in f is a function of $\frac{\partial E_x}{\partial t}$, the variation in amplitude is the first order effect, as would be expected in cases where dissipation overrides linear and nonlinear restoring forces.

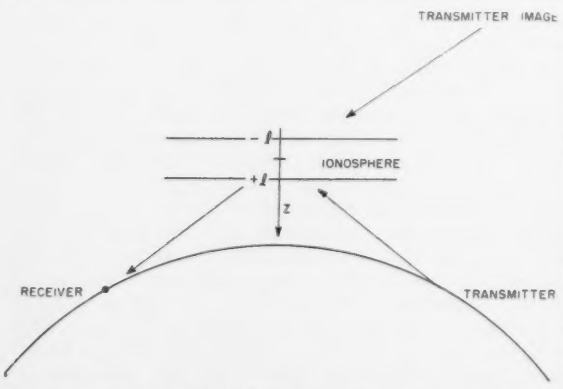


FIGURE 1. *Ionospheric link—downcoming waves generated by horizontally stratified inhomogeneous layer.*

In the case of high frequency radio wave transmission through the ionosphere, it is expected that the cases of interest are those for which the first order effect is a phase distortion. In any event, for direction finding applications, conditions which produce attenuation will reduce the space signal at the receiving antenna and limit the usefulness of this propagation mode. As the radio wave front progresses into the ionospheric layer, some parts of it are advanced in phase with respect to other parts as a result of the blobby, granular index of refraction structure of the medium. It will be assumed that the first order effect of the inhomogeneous layer on the emerging radio field is a phase distortion; which, of course, will produce an amplitude variation in the far field. An optics system is a good example of this phenomenon: a plane light wave incident upon a lens is distorted in phase but not amplitude in a plane immediately behind the lens, and this aperture field distribution creates the usual variation in intensity in the focal plane. Thus, a distortion in the phase of the radio field by the ionosphere will create new space frequencies which modulate the space signal at the earth's surface. The movements of the ionospheric irregularities will create a time and space variation of the phase distortion which in turn produces a time and space variation of the amplitude of the radio field at the surface of the earth.

In order to describe the output of the receiving antenna, it is necessary to specify the space signal in its aperture plane. The phase distortion produced by the ionosphere may be described in terms of its Fourier components. For simplicity, let it be assumed that at $z=l$ the phase distortion has a dominant component along the y direction with amplitude, a, and space frequency or wavenumber k_a. The amplitude of the radio field is assumed constant and the phase in equation (3) is given by

$$\varphi_{z=l}(y)=a \sin k_a y. \tag{13}$$

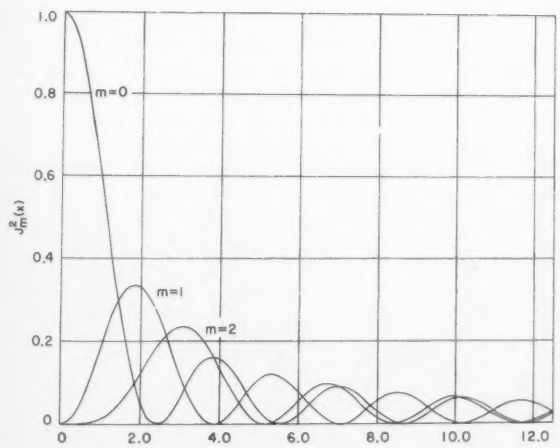


FIGURE 2. *Energy diffracted into first three orders as a function of the radio phase distortion.*

It can be shown [Smyth, 1959] that this phase distortion over a distance $y=2\ b$ will diffract energy into other radio wave space frequencies with amplitude given by

$$A(k_y)=\frac{b}{\pi}\sum_{-\infty}^{\infty} J_m(a)\,\frac{\sin\,(k_y-mk_a)b}{(k_y-mk_a)b}, \tag{14}$$

where $J_m\,(a)$ is the Bessel function of order m, and a, is the amplitude of the phase distortion. The radio space frequency in this case is its value along the y-axis, where, of course, the cos θ in equation (2) is replaced by sin θ. If the region over which the phase distortion takes place is large compared with the radio wave frequency, then the amplitude of the new space frequencies generated by diffraction will differ from zero only in the immediate vicinity of the spectral lines

$$k_{ym}=mk_a \tag{15}$$

since the function

$$\frac{\sin\,(k_y-mk_a)b}{(k_y-mk_a)b} \tag{16}$$

is a maximum value for zero argument. The intensity of the radio field diffracted into each order is given by the square of the amplitude function, $J_m\,(a)$, which is just unity, when $m=0$ and $a=0$. This is the case of no distortion and for this case equation (14) reduces to a single line, $k_y=0$, which is just the original plane wave. It is of interest to look at equation (14) and observe the effect of radio frequency on the amplitude of the diffracted orders. Figure (2) gives $J_m^2\,(a)$ for $m=0,\ 1,\ 2$. It is clear that the energy diffracted out of the original plane wave never completely reappears in any one higher order. For small phase distortion we have

$$\begin{aligned} J_0\left(\frac{2\pi}{\lambda}L\right)&\to 1,\\ J_1\left(\frac{2\pi}{\lambda}L\right)&\to\frac{\pi L}{\lambda},\\ J_2\left(\frac{2\pi}{\lambda}L\right)&\to 1/2\left(\frac{\pi L}{\lambda}\right)^2, \end{aligned} \tag{17}$$

where $a=(2\pi L/\lambda)$ is the amplitude of the radio phase variation from its mean value. Capital L is the deviation of the radio phase from the mean.

3. Antenna as Space Frequency Filter

3.1. Antenna Response

The output voltage of an antenna is determined by its filter characteristics and the electromagnetic field in which it is immersed. For a given orientation of an antenna, the output voltage is a single-valued function of the distribution of electromagnetic field over the effective aperture of the antenna, but unfortunately the converse is not true, there are an infinite number of field distributions that will yield the specified voltage output.

Let the aperture plane of the antenna be the xy plane $z=0$. If the antenna is fed radio energy at a constant time frequency, a resulting field distribution $E(x,y)$ will be maintained over the aperture plane. This aperture distribution may be represented by a spectral distribution of space frequencies in much the same way that a time signal may be represented in the time frequency domain. In terms of plane wave functions, the space representation of the field is given in the space frequency domain by the three dimensional Fourier transform:

$$A(k_x,k_y,k_z)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E(x,y,z)\,e^{+jk_xx+jk_yy+jk_zz}\,dxdydz, \qquad (18)$$

where $A(k_x,k_y,k_z)$ is the amplitude of the plane wave associated with the direction (k_x,k_y,k_z), and, as usual, the space frequency components in the cartesian coordinate system are related as follows:

$$k_x^2+k_y^2+k_z^2=k^2. \qquad (19)$$

For the case where all the space frequency components are real, the fields are propagated waves and the amplitude function $A(k)$ evaluated for the field distribution over any plane is just the polar diagram of this particular radio field distribution [Booker and Clemmow, 1950].

Fields in the xy plane, for which $k_x^2+k_y^2>k^2$ correspond to waves beyond cutoff in the z direction and these make up the so-called "induction field" of the aperture distribution. The distorted radio field emanating from the ionosphere may have components in the current distribution which are of this type, in which case the antenna will be in the induction field for these components. At the distance involved, only the spectral components of the induction field in the immediate region of cutoff will be felt at the ground, since the higher space frequencies will be attenuated rapidly with distance.

Considering the field in the aperture plane $z=0$, the space frequency representation of the radio field and the field distribution in space are related by the two-dimensional Fourier transform pair

$$A(k_x,k_y)=\iint E(x,y)e^{jk_xx+jk_y}dxdy,$$

$$E(x,y)=\iint A(k_x,k_y)e^{-jk_xx-jk_yy}dk_xdk_y. \qquad (20)$$

The output of an antenna placed in this field distribution will simply be the superposition of the responses to the various plane wave components in the field illuminating the antenna aperture. In terms of the space frequencies, the space signal is just

$$A(k_x,k_y)=\iint A(k_x,k_y)\,\delta(k_x-k_x',k_y-k_y')\,dk_x'dk_y', \qquad (21)$$

where δ is the impulse function in the space frequency domain; that is, a plane wave of unit strength arriving from some direction $\bar{k}$ with the property

$$\begin{aligned}\delta(k_x-k_x',k_y-k_y')&=1 \qquad k_x'=k_x\\ &\qquad\quad k_y'=k_y\\ &=0 \qquad k_x'\neq k_x\\ &\qquad\quad k_y'\neq k_y. \end{aligned} \qquad (22)$$

Let the impulse response of the antenna be $G(k_x,k_y)$, then the output will be the sum of the responses created by the input signal and may be written thus:

$$V(k_x,k_y)=\iint G(k_x-k_x',k_y-k_y')A(k_x',k_y')dk_x'dk_y',$$

or,

$$V(k_x,k_y)=\iint G(k_x',k_y')A(k_x-k_x',k_y-k_y')dk_x'dk_y' \qquad (23)$$

since the output must be the same linear combination of the displaced signal as the displaced response. Now, if the signal is the unit impulse function, the output will be just the space frequency filter response function.

$$V(k_x,k_y)=\iint G(k_x',k_y')\delta(k_x-k_x',k_y-k_y')dk_x'dk_y'=G(k_x,k_y). \qquad (24)$$

Since the aperture field input may be expressed as the superposition of impulse functions with appropriate amplitude and phase, the output will be the convolution of the input space frequencies and the impulse response of the antenna, which is the voltage "gain-function" of the antenna. Thus, the antenna output is uniquely given by the input signal; on the other hand, the space signal is not a single-valued function of the output.

It is clear that the output given in equation (24) will be the filter response function when the input is a plane wave function. When the input is of more complex nature, the output will not be simply related to the filter transfer function. For some unknown reason, many radio people have chosen to assume that the input signal is always of the form of a single plane wave, and that the output resulting from the convolution of the actual input signal and the filter response function of the antenna is a new response function for the antenna. This practice of absorbing the unknown aperture distribution of the radio field into the antenna filter function has led to considerable confusion, where some people have gone so far as to state that the antenna placed in a nonplane wave radio field "lost its gain." [4] A more sophisticated way of putting the same thing is the more current statement that there has developed an "antenna-to-medium coupling loss." Of course, the antenna is excited by the radio field

[4] This reminds us of the small boy riding his bicycle down a rough road, concluding that each time a wheel rolled over a depression that the wheel *lost its diameter.*

in its aperture and this radio field is influenced by the medium; but in electromagnetic theory the differences in physical properties of media are accounted for by charge and current distributions. Thus, the only way an antenna can be coupled to a medium is through the fields. For example, it would be of questionable value to couple to a waveguide capable of propagating many modes without regard to the mode being transmitted at the time. Even though the so-called medium remains the same independent of the mode being propagated, the fields are quite different, and the concept of antenna-to-medium coupling loss is of no real utility.

In general, the field distribution across the aperture of the antenna will change with time and consequently the output from the antenna most likely will change with time. Both tropospheric and ionospheric transmission studies indicate that the functional form of the aperture field, $E(x, y,)$, changes with time associated with a continual redistribution of radio energy among the different space frequencies. In direction finding experiments this corresponds to a meandering and spread in the apparent direction and location of the transmitter. Even when the propagation effects are neglected, a time modulation of the radio field will produce an uncertainty in the determination of the angle of arrival. This effect has been discussed by Dausin et al., [1959], who have shown that the error can become appreciable when the time represented by the reciprocal of the bandwidth is less than the propagation time across the antenna aperture.

3.2. Uncertainty Principle

If the space signal could be sampled at a point by an isotropic antenna, and even if the radio field at the point could be measured with unlimited precision, the signal variation in space or direction of arrival would be completely undetermined. To gain information on the space signal which contains information on the arrival direction requires the utilization of space. Again considering the xy plane $z=0$ as the antenna aperture plane, the range of uncertainty in the space frequencies making up the signal along a given direction in the plane is related to the available aperture in that direction by the relationship

$$\Delta|k|\cdot\delta|r|\geq\beta \tag{25}$$

This is, in essence, the celebrated "Uncertainty Principle." The direction property of an antenna is a good example of this principle; if it is desired to radiate most of the energy in a specific direction then the effective aperture must be large. At first glance one might conclude that supergaining violates this relationship, but this is not the case as can be seen from the fact that the effective aperture of the antenna increases with increase in antenna gain according to the relationship (25). Diffraction of light by a slit clearly demonstrates this phenomenon. Any attempt to pass a plane wave through a confined region generates new space frequencies which spread

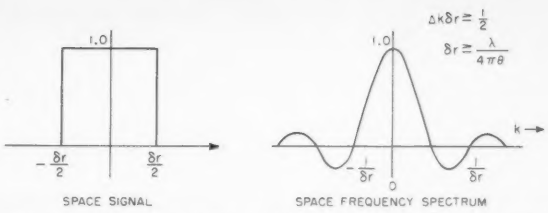


FIGURE 3. *Constant space signal of finite extent and its space frequency spectrum.*

the energy out about the original direction $\bar{k}$ in inverse relationship to the width of the slit.

Suppose that a plane wave is incident normally on the xy plane and its amplitude is measured over a strip of finite width δr. If it is assumed that the space signal is zero outside this range, then the space frequency spectrum is given by

$$C(k)=(\delta r)\,\frac{\sin \pi k(\delta r)}{\pi k(\delta r)}. \tag{26}$$

This space signal and its spectrum is shown in figure 3. On the other hand, any continuous space signal with continuous first derivatives will appear to be constant over a space interval chosen small enough. If, in this case, the space signal is assumed to be constant throughout space, the space frequency spectrum will reduce to a single discrete line. Clearly, then, there are an infinite number of spectral distributions which would yield the same space signal over a given range. To define a spectrum uniquely, the signal must be defined over the entire xy plane. Thus, given a region of space over which the signal will be sampled, (25) places a limit on the accuracy with which the space frequency may be specified.

An example of some practical interest is the space signal distortion generated by the transmission of a plane wave through an inhomogeneous medium such as the earth's atmosphere, including both the troposphere and the ionosphere. This will deteriorate further the accuracy with which the space frequency may be specified and the relationship in (25) is indeed an inequality. The question asked concerns the possibility of filtering out the space frequency noise as a means of reducing the magnitude of the product on the left side of (25). If the characteristics of the space frequency noise are known, it should be possible to design optimum space frequency filters to reduce the uncertainty in the direction of arrival of the desired signal.

4. References

Booker, H. G., and P. C. Clemmow, The concept of an angular spectrum of plane waves and its relation to that of polar diagram and aperture distribution, Proc. IEE **97**, pt. 3, 11–17 (Jan. 1950).

Dausin, L. R., K. E. Niebuhr, and J. J. Nilsson, The effects of wide-band signals on radar antenna design, IRE WESCON Conv. Record (Aug. 1959).

Ratcliffe, J. A., Some aspects of diffraction theory and their application to the ionosphere, Repts. Progr. Phys. **19**, 188–266 (1956).

Smyth, John B., Radio phase distortion—A mechanism for transhorizon radio propagation, URSI Washington meeting (May 1959).

(Paper 65D3–131)

JOURNAL OF RESEARCH of the National Bureau of Standards—Radio Propagation
Vol. 65D, No. 3, May–June 1961

Effect of Receiver Bandwidth on the Amplitude Distribution of VLF Atmospheric Noise [1, 2]

Forrest F. Fulton, Jr.[3]

(October 27, 1960; revised November 8, 1960)

The distribution function of envelope voltage for short samples of atmospheric radio noise as received by a communications receiver in the VLF range always shows a marked departure from that obtained for Gaussian noise. If it is considered that this departure is caused by strong noise pulses which do not overlap in time, the effect of changes in the receiver bandwidth on the observed distribution function can be deduced by consideration of the changes in the receiver impulse response. A transformation can be obtained which gives an excellent approximation to the change in a mathematical representation of the distribution function in the range of probabilities below 1 percent. Empirical relationships are suggested which give useful estimates of the change in the distribution function over the total range of probabilities.

1. Introduction

In the VLF range, atmospheric radio noise is one of the important factors in a system design. If all of the manmade interference is controlled by allocation procedures and good engineering, the atmospheric noise provides the ultimate background disturbance from which the desired signal must be separated [CCIR Rpt. 65, ITU, 1957; Watt, Coon, Maxwell, and Plush, 1958]. Studies of the noise must be of a statistical nature, but there are problems which differ, for instance, from those of statistical studies of thermal noise, because the atmospheric noise is a nonstationary process. Since the statistics of the process change with time, the accuracy of measuring statistical parameters cannot be indefinitely increased by increasing the length of time of the measurements. Experience has shown that when measuring the amplitude probability distribution of the noise envelope, samples of noise 10 to 20 min in length are short enough to avoid difficulties due to the nonstationary characteristics of the process, but are long enough to give useful information for system design purposes [Hoff and Johnson, 1952; Watt and Maxwell, 1957].

One of the problems which occurs in system design is determining the percentage of time that the envelope of the noise will exceed the signal level, and how this varies as the receiver bandwidth is changed. In principle this can be calculated precisely if enough is known about the statistics of the noise. An applicable procedure, as described by Widrow [1957], is to consider a sampled version of the input noise for which the joint probability distribution of the samples can be determined, and to use this to calculate the probability distribution of a sampled version of the filter output. If the sampling rate at the input is high enough, the probability distribution of the output samples will accurately represent the probability distribution of the continuous output. The difficulty with this procedure is that the interval between samples at the input must in general be short relative to the correlation time of the noise [Ragazzini and Franklin, 1958], so that a very high order joint probability density function must be calculated for the filter output. The purpose of this paper is to present a technique which requires much less computation, but gives an accuracy commensurate with the statistical knowledge of the noise.

2. Noise Characteristics

Figure 1 shows two samples of measurements of atmospheric noise envelope distributions, one of relatively high dynamic range conditions and the other of noise of relatively low dynamic range; the dotted

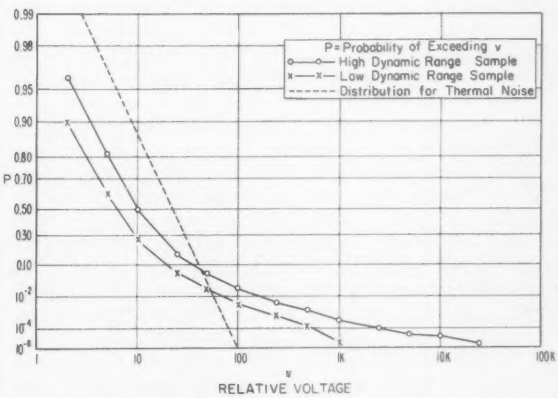


FIGURE 1. *Typical measurements of atmospheric noise envelope distribution.*

[1] Contribution from Central Radio Propagation Laboratory, National Bureau of Standards, Boulder, Colo.
[2] An earlier version of this paper was presented at the Symposium on VLF Radio Waves held in Boulder, Colo., January 1957.
[3] Present address: Lockheed Aircraft Corporation, Palo Alto, Calif.

line represents the measurement which is obtained for thermal noise. At the low voltages which have a high probability of being exceeded, the atmospheric noise measurements parallel those for thermal noise; this is characteristic of a phenomenon consisting of a large number of overlapping pulses, no one of which contributes a significant proportion of the total energy. At the high voltages, however, which have a low probability of being exceeded, the measurements depart radically from the shape of the thermal noise curve, in the form of a much higher probability of exceeding the voltage levels in the range of probabilities of 0.3 and below. Oscillograms of atmospheric noise show that the high voltage levels are attained by occasional noise pulses which greatly exceed the general level of the noise [Yuhara, Ishida, and Higashimura, 1956] and that the pulse shape is that of the impulse response of the receiver. These strong pulses occur infrequently and two such pulses very rarely overlap.

3. Probabilities Associated With Impulses

The effect of bandwidth changes on the probabilities associated with noise impulses can be calculated by considering the effect of bandwidth changes on the usual impulse response of the filter. The effect of increasing the bandwidth is to cause the impulse response to become shorter in time, and higher in amplitude [Guillemin, 1956]. More specifically, if the network characteristics are changed in such a way that the shape of the impulse response is preserved while the bandwidth is multiplied by a factor B, the voltage scale of the impulse response will be multiplied by a factor B and the time scale by a factor $1/B$, as illustrated in figure 2 for $B=2$. This relationship also holds for the envelope of the response of a band-pass circuit whenever the circuit possesses a low pass analog, as is normally the case with receiver circuits [Aigrain, Teare, and Williams, 1949; Guillemin, 1953]. Further, the probabilities associated with the impulse response are determined by the time scale; for example, the probability that a voltage level of one will be exceeded by the narrow band response is determined by the ratio of the time τ to the total observation time. Because of the relationship between changes in the voltage and time scales, the doubled bandwidth impulse response exceeds a voltage level of two for a time of $\tau/2$, which represents precisely one-half of the probability corresponding to a time τ.

Thus the amplitude distribution of a short sample of noise which contained only one impulse would be determined by the time and voltage scales of that impulse response; if a particular voltage v is measured to have a probability P of being exceeded when a single impulse is observed through a network of bandwidth $\Delta\omega_1$, and simultaneously the impulse is observed through a network of bandwidth $\Delta\omega_2 = B\Delta\omega_1$, the voltage Bv will be measured to have a probability $(1/B)\ P$ of being exceeded. This relationship also holds for a noise sample which consists of a number

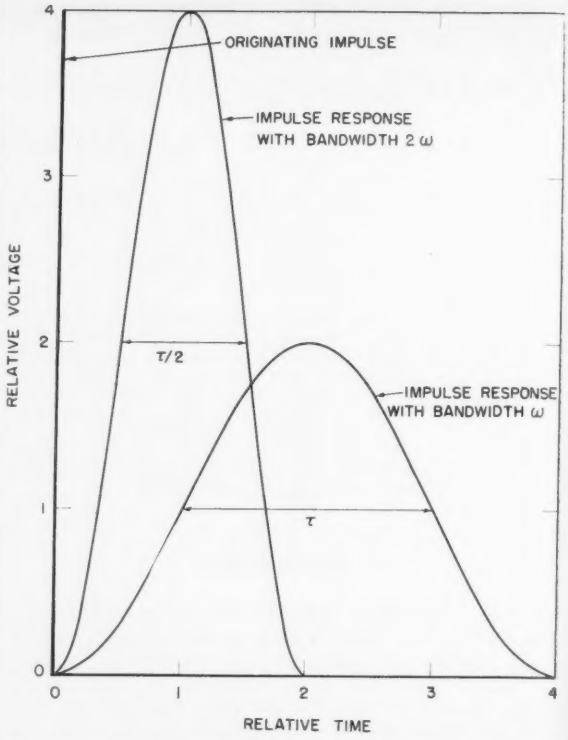


FIGURE 2. *Envelope of band-pass impulse response for two bandwidths.*

of impulses, provided that the response of the network to one impulse does not overlap the response to another.

These requirements are never exactly met in practice because of the exponential decay of physical networks and because of the presence of thoroughly overlapping low level noise. However, the high voltage end of the distribution of VLF atmospheric noise appears to represent an excellent approximation to the requirements. Figure 3 shows the high voltage portion of two representative sets of measurements at two bandwidths and also shows the results of translating the narrow-band measurements by the ratio of bandwidths to expected measurements at the wider bandwidth. It is evident that for both the low and high dynamic range noise conditions the points calculated from the narrow-band measurements agree with the measurements at the wider bandwidth for probabilities less than 0.01.

4. Mathematical Representation of the Noise

In dealing with a non-Gaussian random variable, it has proved profitable for some purposes to consider it as being generated as the output of a nonlinear resistance network which has a Gaussian input, even though the actual phenomenon may be very much

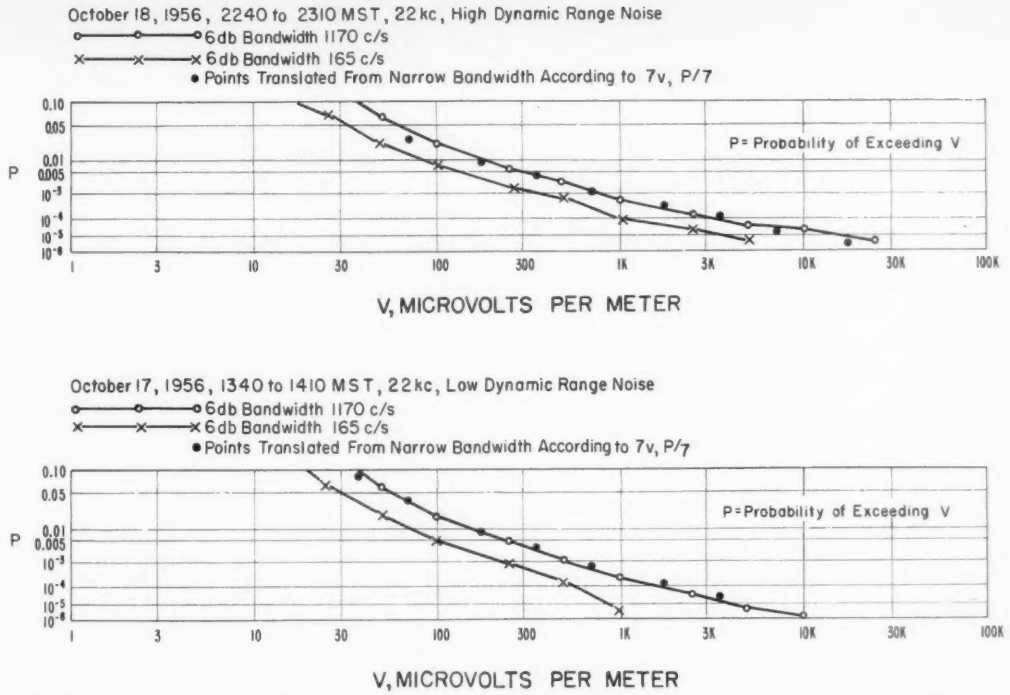


FIGURE 3. *Atmospheric noise distributions showing effect of translating measurements made at a narrow bandwidth to expected values at a wider bandwidth.*

different from this [Smith, 1959a and b]. In the situation being considered here where the envelope of the noise is the quantity of interest, we can represent the envelope of the actual noise as being obtained from an envelope detector having a Gaussian input, followed by a nonlinear resistance network. The justification for doing this is simply that the resulting probability distribution can represent the probability distribution of the actual phenomenon quite accurately, and the model proves to be a convenient one for the type of mathematical manipulations which are required.

A nonlinear network which can produce the correct probability distribution has a voltage transfer characteristic which is a three term polynomial of the form

$$v=a_1y+a_2y^{\frac{b+1}{2}}+a_3y^b \qquad (1)$$

where a_1, a_2, a_3, and b are chosen to match the atmospheric noise distribution under consideration; the probability that a voltage level v is exceeded at the output of the network is given by

$$P=e^{-y^2}. \qquad (2)$$

The variable y may conveniently be considered just as a parametric variable, although it is the envelope voltage of a Gaussian variable having a variance of one-half.

Each term of the polynomial, if considered separately, plots as a straight line of log-log of probability versus logarithm of voltage, which are the coordinate scales used on the graphs of distributions in figures 1 and 3. The first term represents a distribution having complete overlapping, such as thermal noise; the second and third terms represent departures from this distribution. The third term of the polynomial, a_3y^b, is the dominant term at the high voltage, low probability end of the curve, when the phenomena forming the distribution appear to be strong, nonoverlapping pulses.

5. The Effect of Bandwidth Changes on the Mathematical Model

Since the third term of the polynomial is considered to represent nonoverlapping impulses, the effect of bandwidth changes can be obtained by the procedure described in section 3. That is, for an increase in bandwidth by a factor B, any point on the original line at voltage v and probability P is transformed to a voltage Bv at a probability P/B. After this is done, it is found that the points no longer lie on a straight line, but along a line having a slight curvature, as illustrated in figure 4 for an exponent of 8 and a bandwidth increase of ten times. The curvature introduced in this way is so slight, however, that the curve may be replaced by a line tangent to

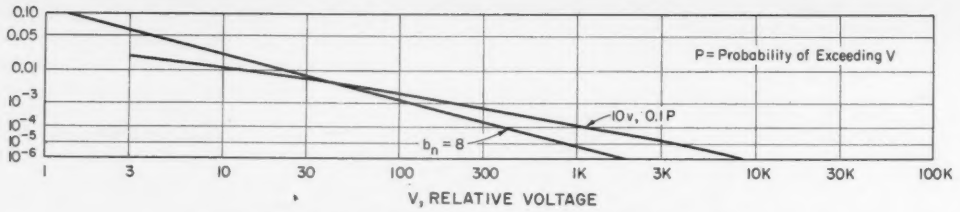


FIGURE 4. *Curvature introduced by bandwidth transformation.*

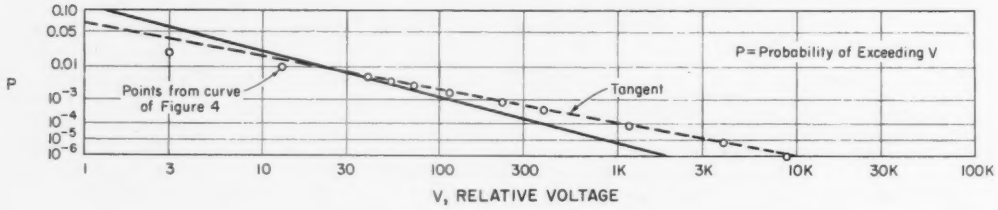


FIGURE 5. *Tangent approximation to curve.*

it at the point where $P=e^{-8}$; as is apparent from figure 5, this is an excellent approximation.

It is possible to derive a general formula for the effects of bandwidth changes on the exponent and the coefficient of the third term of the polynomial. For a bandwidth increase by a factor of B, the exponent b_w for the wider bandwidth is obtained from the exponent b_n for the narrow bandwidth by the relationship

$$b_w = \frac{8}{8 - L_n B}\, b_n \tag{3}$$

where $L_n B$ is the natural logarithm of B. The coefficient is obtained from

$$a_{3w} = a_{3n} B \left[(8 - L_n B) 8^{\left(\frac{-8}{8 - L_n B}\right)} \right]^{\frac{b_n}{2}} \tag{4}$$

where a_{3w} and a_{3n} are the coefficients for the wide and narrow bandwidths respectively. The derivation of these relationships is given in the appendix.

For the other terms of the probability function the effects of bandwidth changes are not as apparent as for the third term which describes the behavior at high voltages. The second term of the polynomial is important in the moderate voltage range where some overlapping of noise pulses occurs. For the values of exponents and coefficients which are required to match the mathematical model to the observed atmospheric noise conditions, it happens that the second moment of the distribution is almost entirely determined by the second and third terms of the polynomial. Since for normal communications bandwidths the total noise power received is directly proportional to the bandwidth, it is possible to adjust the second term to obtain the correct behavior for the second moment of the distribution. The procedure is to calculate the changes in the a_3 coefficient and exponent b as described above, and to insert these into the expression for the mean squared voltage of this form of distribution, which is

$$\langle v^2 \rangle = a_1^2 + a_2^2 \Gamma\left(\frac{b+3}{2}\right) + a_3^2 \Gamma(b+1) + 2a_1 a_2 \Gamma\left(\frac{b+7}{4}\right) + 2a_1 a_3 \Gamma\left(\frac{b+3}{2}\right) + 2a_2 a_3 \Gamma\left(\frac{3b+5}{4}\right) \tag{5}$$

where $\langle v^2 \rangle$ is the mean squared voltage.

For a bandwidth increase by a factor of B, the mean squared voltage for the wide bandwidth is B times the mean squared voltage at the narrow bandwidth. With this and a value for a_1, equation (5) can then be solved for the coefficient a_2.

The first term of the polynomial, involving a_1, represents a situation where complete overlapping occurs. The slope of a distribution represented by only this term does not change with bandwidth, but only its amplitude, as long as the complete overlapping is maintained. If this were the only term in the distribution, its amplitude would change precisely as the square root of the bandwidth. Since there are other terms present, however, there will be an interchange of energy between the various portions of the distribution, and the net effect will be that the amplitude of the first term will change more slowly than this; an increase according to the four-tenths power of the bandwidth appears to fit the available experimental data quite well.

These procedures work quite well with noise samples of low and moderate dynamic range, but an error appears when used with noise of high dynamic range, that is, values of b of about 11 or greater. This occurs because the mathematical

model being used is not sufficiently complex. With the very high dynamic range conditions, a substantial contribution to the integral for the mean squared voltage occurs at probabilities below 10^{-6}; in this region the tangent line and the curve representing the distribution diverge significantly, causing considerable error in the evaluation of the mean squared voltage.

Under the high dynamic range conditions it seems appropriate to make the same transformation on the second term as on the third. Modification of the coefficient by some factor to compensate for the partial overlapping might be considered, but also considering the nature of the experimental data, that is, short samples of a nonstationary process, and also the relative unimportance of the second term, any refinement in the transformation seems unjustified.

6. Discussion

These somewhat arbitrary operations on the first and second terms, together with the accurate transformation for the high voltage end of the distribution where the third term is dominant, provide a method of describing what changes would be observed in the polynomial representation of the distribution of atmospheric radio noise amplitude when observed through different bandwidths. The results of carrying out the complete transformation on representative samples of high and low dynamic range noise are shown in figure 6 and figure 7. For the low dynamic range noise in figure 6, the a_2 coefficient was determined from the rms voltage relationship; for the high dynamic range noise in figure 7, it was determined by the transformation procedure for nonoverlapping pulses. It is apparent that these procedures provide good estimates of the distributions, and are useful and practical tools for use when more rigorous methods are not justified.

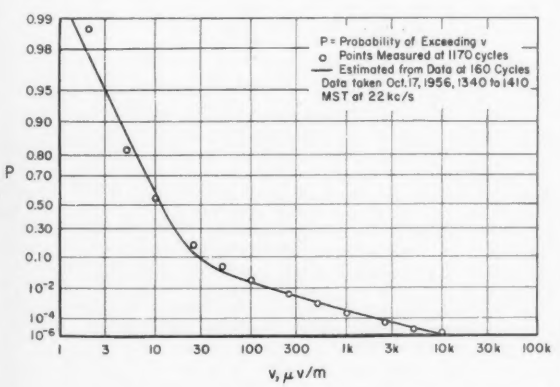


FIGURE 6. *Low dynamic range atmospheric noise distribution measured at a wide bandwidth, and estimated distribution from simultaneous narrow bandwidth data.*

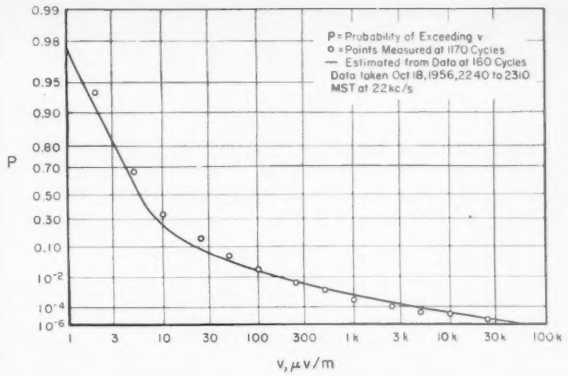


FIGURE 7. *High dynamic range atmospheric noise distribution measured at a wide bandwidth, and estimated distribution from simultaneous narrow bandwidth data.*

I wish to acknowledge the experimental data made available by A. D. Watt and E. L. Maxwell. L. P. Benedict provided valuable assistance with much of the numerical work.

7. Appendix: Variation of High Voltage Coefficient and Exponent

Consider a distribution having a form expressed by

$$v_1 = ay^b$$

$$P_1 = e^{-y^2} \tag{6}$$

when it is observed through a particular bandwidth. If this distribution is composed of nonoverlapping impulses, then when observed through a bandwidth B times larger, the voltages and probabilities will be related by the ratio of bandwidths according to

$$P_2 = \frac{P_1}{B}$$

$$v_2 = Bv_1 \tag{7}$$

where P_1 and v_1 are obtained at the first bandwidth and P_2 and v_2 at the second. From (6) and (7) we can write

$$\log(-LnP_1) = \frac{2}{b}\log\left(\frac{v_1}{a}\right)$$

$$\log(-LnBP_2) = \frac{2}{b}\log\left(\frac{v_2}{aB}\right) \tag{8}$$

We wish to develop an approximation to the relationship between v_2 and P_2 in the region of high voltages and low probabilities which will have the form

$$v_2 = aBC_1 y^{bC_2}$$

$$P_2 = e^{-y^2} \tag{9}$$

or

$$\log(-LnP_2)=\frac{2}{bC_2}\log\left(\frac{v_2}{aBC_1}\right) \tag{10}$$

where C_1 and C_2 are constants to be determined. They may be evaluated by selecting two points on the P_2 versus v_2 relationship where the approximation is to be exact. Choosing the points $P_{2\alpha}$, $v_{2\alpha}$, and $P_{2\beta}$, $v_{2\beta}$, we obtain from (8)

$$v_{2\alpha}=aB(-LnBP_{2\alpha})^{\frac{b}{2}}$$

$$v_{2\beta}=aB(-LnBP_{2\beta})^{\frac{b}{2}}. \tag{11}$$

Now in the approximation,

$$\log(-LnP_{2\alpha})=\frac{2}{bC_2}\log\left(\frac{v_{2\alpha}}{aBC_1}\right)$$

$$\log(-LnP_{2\beta})=\frac{2}{bC_2}\log\left(\frac{v_{2\beta}}{aBC_1}\right) \tag{12}$$

C_1 may be eliminated from equations (12), obtaining

$$\log\left(\frac{LnP_{2\alpha}}{LnP_{2\beta}}\right)=\frac{2}{C_2b}\log\left(\frac{v_{2\alpha}}{v_{2\beta}}\right) \tag{13}$$

and from (11),

$$\frac{v_{2\alpha}}{v_{2\beta}}=\left(\frac{LnBP_{2\alpha}}{LnBP_{2\beta}}\right)^{\frac{b}{2}} \tag{14}$$

and from (13) and (14) we can obtain

$$C_2=\frac{\log\left(\frac{LnBP_{2\alpha}}{LnBP_{2\beta}}\right)}{\log\left(\frac{LnP_{2\alpha}}{LnP_{2\beta}}\right)}. \tag{15}$$

For the evaluation of C_1, from (11) and (12) we can obtain

$$C_1=\frac{(-LnBP_{2\alpha})^{\frac{b}{2}}}{(-LnP_{2\alpha})^{\frac{bC_2}{2}}}=\frac{(-LnBP_{2\beta})^{\frac{b}{2}}}{(-LnP_{2\beta})^{\frac{bC_2}{2}}}. \tag{16}$$

Going back to the polynomial term itself, we can change the third term to read $a_{3w}y^{b_w}$, where

$$b_w=b_nC_2$$

and

$$a_{3w}=a_{3n}BC_1 \tag{17}$$

a_{3n} and b_n being the values observed at the narrow bandwidth.

A limiting value of C_2 is obtained as $P_{2\alpha}$ approaches $P_{2\beta}$ and this represents the approximation where the straight line is tangent to the curve. This limiting value may be found by application of l'Hospital's rule, and is

$$C_2=\frac{LnP_{2\beta}}{LnBP_{2\beta}}. \tag{18}$$

If we use the tangent approximation and evaluate for

$$P_{2\alpha}=P_{2\beta}=e^{-8}=0.000335$$

we obtain

$$b_w=C_2b_n=\frac{8}{8-LnB}\,b_n \tag{19}$$

and

$$a_{3w}=a_{3n}B\left(\frac{8-LnB}{8C_2}\right)^{\frac{b_n}{2}}. \tag{20}$$

Figure 5 shows this tangent line for an assumed b_n of 8 and B of 10, and also shows some points from the curve in figure 4. The difference between the straight line and the points from the curve is zero, of course, at $P=0.000335$, and increases slowly to about 1 db at $P=0.005$ and $P=10^{-6}$; it is evident that there is no significant difference between using the straight line and the curve, and the procedure using the tangent line provides a simple and accurate estimate of the changes in the high voltage end of the distribution curve.

8. References

Aigrain, P. R., B. R. Teare, Jr., and E. M. Williams, Generalized theory of the band-pass low-pass analogy, Proc. IRE **37,** 1152–1155 (1949).

Consultative Committee on International Radio, Revision of atmospheric radio noise data, CCIR Rpt. 65, International Telecommunication Union, Geneva (1957).

Guillemin, E. A., Introductory circuit theory, p. 434 (John Wiley and Sons, Inc., New York, N.Y., 1953).

Guillemin, E. A., The mathematics of circuit analysis, p. 526 (John Wiley and Sons, Inc., New York, N.Y., 1956).

Hoff, R. S., and R. C. Johnson, A statistical approach to the measurement of atmospheric noise, Proc. IRE **40,** 185 (1952).

Ragazzini, J. R., and G. F. Franklin, Sampled-data control systems, p. 261 (McGraw-Hill Book Co., Inc., New York, N.Y., 1958).

Smith, O. J. M., Spectral output of piecewise linear nonlinearity, A.I.E.E. Transactions, pt. 1, **78,** 543 (1959).

Smith, O. J. M., Statistical spectral output of power law nonlinearity, A.I.E.E. Transactions, pt. 1, **78,** 535 (1959).

Watt, A. D., and E. L. Maxwell, Measured statistical characteristics of VLF atmospheric radio noise, Proc. IRE **45,** 55 (1957).

Watt, A. D., R. M. Coon, E. L. Maxwell, and R. W. Plush, Performance of some radio systems in the presence of thermal and atmospheric noise, Proc. IRE **46,** 1914 (1958).

Widrow, B., Propagation of statistics in systems, IRE Wescon Convention Record, pt. 2, 114 (1957).

Yuhara, H., T. Ishida, and M. Higashimura, Measurement of the amplitude probability distribution of atmospheric noise, Journal of the Radio Research Laboratories, Tokyo, Japan, **3,** 101 (1956).

(Paper 65D3–132)

JOURNAL OF RESEARCH of the National Bureau of Standards—D. Radio Propagation
Vol. 65D, No. 3, May–June 1961

Excitation of VLF and ELF Radio Waves by a Horizontal Magnetic Dipole [1, 2]

Janis Galejs

(August 29, 1960; revised November 10, 1960)

The VLF and ELF modes excited by a horizontal magnetic dipole (vertical loop) in the spherical shell between a finitely conducting earth and an isotropic sharply bounded ionosphere are shown to have a nearly transverse magnetic character. The modes are similar to those of a vertical electric dipole. With the exception of the zero order mode, the propagating modes excited by the magnetic dipole are of slightly higher amplitudes, provided that the far fields of the horizontal magnetic and vertical electric dipoles are equal over flat earth in the absence of ionosphere.

The transient fields generated by a current step in the magnetic dipole are in the first approximation similar to the fields generated by a current impulse in a vertical electric dipole. Response of the zero order mode of the magnetic dipole has been calculated.

1. Introduction

The mode theory of VLF transmissions has been developed by Watson [1919], Budden [1953], Schumann [1954], and Wait [1957, 1960a, 1960b] for excitation by vertical and horizontal electric dipoles and also by vertical magnetic dipoles (horizontal loops). Additional references on this subject have been listed by Wait [1960a, 1960b].

The reciprocity theorem has been found to be useful for relating the fields of a horizontal electric dipole to fields of vertical electric and magnetic dipoles [Wait, 1960a]. The reciprocity relations may be also applied to relate the fields due to horizontal magnetic dipole excitation to known fields due to vertical electric and magnetic dipoles.

2. Reciprocity Relations

The reciprocity theorem states that the voltage V_2 induced in antenna 2 by current I_1 of antenna 1 is the same as the voltage V_1 induced in antenna 1 by an identical current I_2 flowing in antenna 2.

A vertical electric (VE) dipole of length ds at $z_0'=z_t$ and a horizontal magnetic (HM) dipole parallel to the x-axis (or a vertical loop of area da in the y, z plane) at $z_0=z_r$ are shown in figure 1, where subscripts r and t refer to receiver and transmitter coordinates respectively. The only nonzero H component generated by the VE dipole is $H_{\phi'}$. The magnitude of the voltage induced in the HM dipole V^{hm} is maximum if $H_{\phi'}$ is parallel to the x-axis, which occurs for $\phi=\frac{\pi}{2}$ and $\frac{3\pi}{2}$. For $e^{i\omega t}$ variation of the fields.

$$V^{hm}=-i\omega\mu_0 H_{\phi'}^{ve}(z_0'=z_t,\ z_0=z_r)\sin\phi\ da, \qquad (1)$$

where the superscripts hm and ve refer to horizontal magnetic and vertical electric dipoles respectively. With the same current applied to the HM dipole only the vertical electric field E_z will contribute to V^{ve} and

$$V^{ve}=E_z^{hm}(z_0=z_t, z_0'=z_r)ds. \qquad (2)$$

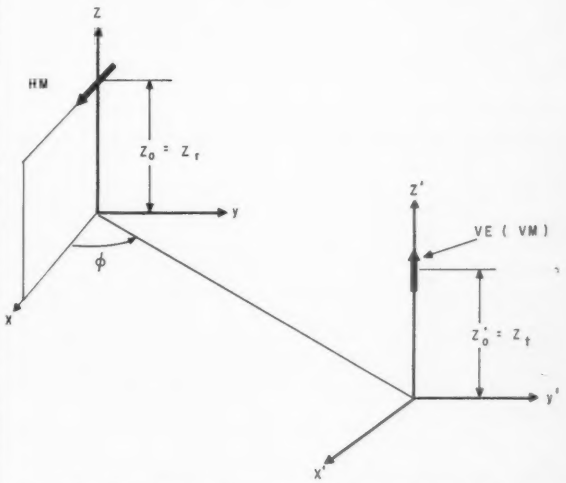


FIGURE 1. *Coordinate systems for defining reciprocity relations.*

[1] Contribution from Applied Research Laboratory of Sylvania Electronic Systems, Waltham, Mass.

[2] The research reported in this paper was supported by a contract of the Office of Naval Research.

Equating (1) and (2) results in

$$E_z^{hm}(z_0=z_t,z_0'=z_r)$$

$$=-i\omega\mu_0 \sin\phi H_{\phi'}^{ve}(z_0'=z_t,z_0=z_r)da/ds. \quad (3)$$

A vertical magnetic (VM) dipole (or a horizontal loop of area da^{vm} parallel to the x', y' plane) at $z_0'=z_t$ and a horizontal magnetic (HM) dipole parallel to the x-axis (or a vertical loop of area da^{hm} in the y, z plane) at $z_0=z_r$ may be considered in the geometry shown in figure 1. The only nonzero H component generated by the VM dipole is $H_{\rho'}$. The magnitude of the voltage induced in the HM dipole V^{hm} is maximum, if $H_{\rho'}$ is parallel to the x-axis, which occurs for $\phi=0$ or π. Hence

$$V^{hm}=i\omega\mu_0 H_{\rho'}^{vm}(z_0'=z_t,\ z_0=z_r)\cos\phi da^{hm} \quad (4)$$

where the superscript vm refers to the vertical magnetic dipole. With the same current applied to the HM dipole only the vertical magnetic field H_z will contribute to V^{vm} and

$$V^{vm}=-i\omega\mu_0 H_z^{hm}(z_0=z_t,z_0'=z_r)da^{vm}. \quad (5)$$

Equating (4) and (5) gives for $da^{vm}=da^{hm}$

$$H_z^{hm}(z_0=z_t,z_0'=z_r)=-\cos\phi H_{\rho'}^{vm}(z_0'=z_t,z_0=z_r). \quad (6)$$

For spherical coordinates the z component is replaced by r component and the ρ component by the θ component of the fields.

The reciprocity relations apply to dipoles in the presence of any linear media. The fields of dipoles above plane earth derived by solving the wave equation should also satisfy (3) and (6). Thus (27) and (116) of Norton [1937], which have been derived for unit dipoles, check the validity of (3). Some algebra is involved in verifying (6) from (35.1) of Sommerfeld [1949] and from (2) to (6), (17) to (19), and (30) of Norton [1937].

3. Fields of the Horizontal Magnetic Dipole

The dipole fields will be examined with the aid of reciprocity relations in an idealized spherical shell that is bounded by a homogeneous earth of radius a, conductivity σ_g and dielectric constant ϵ_g and by a homogeneous ionosphere of radius $(a+h)$, conductivity σ_i and dielectric constant ϵ_i. The above model neglects the effects of the earth's magnetic field.

The vertical electric field E_r^{hm} is computed by (3) applying (5.1), (6.16) of Wait [1960a] and using the relation

$$\frac{d}{d\theta}P_\nu(-\cos\theta)$$

$$=\nu[\cos\theta P_\nu(-\cos\theta)+P_{\nu-1}(-\cos\theta)]/\sin\theta. \quad (7)$$

This results in

$$E_r^{hm}=\frac{I(da)i\omega\mu_0 \sin\phi}{2hr_r}\sum_{n=0}^{\infty} f_n(z_t)f_n(z_r)\cdot\frac{\delta_n}{\sin\nu\pi}\frac{\nu}{\sin\theta}$$

$$[\cos\theta P_\nu(-\cos\theta)+P_{\nu-1}(-\cos\theta)]. \quad (8)$$

The vertical magnetic field H_r^{hm} is computed by (6) applying (9.1), (9.2) of Wait [1960a] and using the relation (7) as

$$H_r^{hm}=-\frac{I(da)\omega\mu_0\cos\phi}{2khr_t r_r\eta}\sum_{m=1}^{\infty}\frac{\delta_m f_m(z_r)\mu}{\sin\mu\pi\sin\theta}$$

$$\left[f_m(z_t)+r_t\frac{d}{dz_t}f_m(z_t)\right]$$

$$[\cos\theta P_\mu(-\cos\theta)+P_{\mu-1}(-\cos\theta)]. \quad (9)$$

Using the second order representation of the radial functions, applying (6.9), (6.17), (9.3), and (9.4) of Wait [1960a] yields

$$2f_q(z)=e^{ikC_q z}[R_g^q(C_q)]^{-1/2}+e^{-ikC_q z}[R_g^q(C_q)]^{1/2} \quad (10)$$

$$\delta_q=\left(1\pm\frac{\sin 2khC_q}{2khC_q}\right)^{-1} \quad (11)$$

where q is an integer equal to either n or m and where

$$R_p^n(C_n)=\frac{n_pC_n-C_p}{n_pC_n+C_p} \quad (12)$$

$$R_p^m(C_m)=\frac{C_m-n_pC_p}{C_m+n_pC_p} \quad (13)$$

$$C_q=\sqrt{1-S_q^2}, \quad (14)$$

$$C_p=\sqrt{1-(S_q/n_p)^2}, \quad (15)$$

$$S_q=\frac{\rho+0.5}{ka}. \quad (16)$$

The plus sign of (11) should be used with $q=n$, while the minus sign is appropriate for $q=m$. In (16) ρ is a complex number, which is equal to ν of (8) or equal to μ of (9), $k=\omega/c$, c=velocity of light. The refractive index n of the boundary medium characterized by subscript p is

$$n_p^2=(\sigma_p+i\omega\epsilon_p)/\epsilon_0 \quad (17)$$

where σ_p and ϵ_p are the conductivity and permittivity of the boundary medium and where ϵ_0 is the permittivity of free space ($\epsilon_0=(36\pi)^{-1}\ 10^{-9}$ farad/meter). The boundary medium can be either ground ($p=g$) or ionosphere ($p=i$). The refractive index of ground is

$$n_g^2\approx\sigma_g/(i\omega\epsilon_0) \quad (18)$$

where σ_g is the ground conductivity. The refractive index of the ionosphere is

$$n_i^2=1+\sigma_i/(i\omega\epsilon_0). \tag{19}$$

The ionospheric conductivity σ_i is defined by

$$\sigma_i=\frac{\epsilon_0\omega_p^2}{\nu} \tag{20}$$

where ν is the electron collision frequency in the ionosphere and ω_p is the plasma frequency ($\omega_p^2=3180\ N$, N=number of electrons per meter^3). C_n and C_m are roots of the modal equations

$$R_i^q(C_q)R_g^q(C_q)=\exp(+2ikhC_q). \tag{21}$$

S_q that is related to C_q by (14) has a magnitude of approximately unity and a small imaginary part for propagating modes of low attenuation. With S_n and S_m determined from (14), ν of (8) and μ of (9) follow from (16). This completes the formal specification of the fields E_r^{hm} and H_r^{hm}.

The expressions for E_r^{hm} and H_r^{hm} may be simplified by introducing further approximations. The consideration will be restricted to cases where

$$r_r\approx r_t\approx a, \tag{22}$$

$$kC_q z_r\ll 1, \tag{23}$$

$$kC_q z_t\ll 1, \tag{24}$$

$$n_g\gg 1. \tag{25}$$

This gives

$$R_g^n(C_n)\approx 1-\frac{2}{n_g C_n}\approx 1, \tag{26}$$

$$R_g^m(C_m)\approx -1+\frac{2C_m}{n_g C_g}\approx -1, \tag{27}$$

$$f_n(z)\approx 1+\frac{1}{8n_g^2C_n^2}\approx 1, \tag{28}$$

$$f_m(z)\approx C_m/(in_gC_g)\approx C_m/(in_g), \tag{29}$$

$$df_m(z)/dz=ikn_gC_gf_m(z)\approx ikn_gf_m(z). \tag{30}$$

Approximating the Legendre functions by [Watson, 1919; Bremmer, 1949]

$$P_\rho(-\cos\theta)\approx(2\pi\rho\sin\theta)^{-0.5}\exp\left[i\left(\rho+\frac{1}{2}\right)(\pi-\theta)-i\pi/4\right] \tag{31}$$

results in

$$E_r^{hm}=-\sin\phi E_0^{hm}\sqrt{\frac{d/a}{\sin(d/a)}}\sqrt{\frac{d/\lambda}{h/\lambda}}\,e^{i\frac{2\pi d}{\lambda}-i\frac{\pi}{4}}\cdot\sum_{n=0}^{\infty}\delta_n S_n^{0.5}e^{-i\frac{2\pi d}{\lambda}S_n}, \tag{32}$$

$$H_r^{hm}=-\cos\phi\frac{E_0^{hm}}{\eta n_g}\sqrt{\frac{d/a}{\sin(d/a)}}\sqrt{\frac{d/\lambda}{h/\lambda}}\,e^{i\frac{2\pi d}{\lambda}-i\frac{\pi}{4}}\cdot\sum_{m=1}^{m}\delta_m(1-S_m^2)S_m^{0.5}e^{-i\frac{2\pi d}{\lambda}S_m}, \tag{33}$$

where the distance along the curved earth is

$$d=a\theta, \tag{34}$$

and where

$$E_0^{hm}=\frac{2\pi I(da)\eta}{\lambda^2 d}\,e^{-i\frac{2\pi d}{\lambda}} \tag{35}$$

is the vertical electric field of the source at a distance d in the direction of maximum intensity on a perrectly conducting plane earth.

4. Comparison of Harmonically Excited Dipole Fields

The fields of the horizontal magnetic dipole E_r^{hm} and H_r^{hm} will be compared first with the corresponding field components of the horizontal electric dipole. It follows from (9.37) and (9.42) of Wait [1960a] that for the n^{th} and the m^{th} mode

$$\left.\frac{E_r^{hm}}{E_0^{hm}}\right|_n=-n_g\tan\phi\left.\frac{E_r^{he}}{E_0^{ve}}\right|_n \tag{36}$$

$$\left.\frac{H_r^{hm}}{E_0^{hm}}\right|_m=n_g\cot\phi\left.\frac{H_r^{he}}{E_0^{ve}}\right|_m \tag{37}$$

where the superscripts *he* and *ve* designate field components of horizontal electric and of vertical electric dipoles respectively. E_0^{hm} is defined in (35), E_0^{ve} is the corresponding expression of the vertical electric dipole defined by (6.25) or (9.38) of Wait [1960a]. The approximation $C_g\approx 1$ was used in the expressions for E_r^{he} and H_r^{he} in order to make them consistent with the derivation leading to (32) and (33). $C_g\approx 1$ constitutes a better approximation than $f_n(z)\approx 1$ and the assumption $f_n(z)\approx 1$ [Wait, 1960a] should be followed by $C_g\approx 1$.

Comparison of (32) and (33) shows that ηH_r^{hm} is several orders of magnitude smaller than E_r^{hm} because

$$n_g^{-1}\ll 1 \tag{38}$$

and because

$$1-S_m^2\ll 1 \tag{39}$$

for propagating modes of low attenuation. The H_r^{hm} fields of a given mode are proportional to $n_g^{-1}\sim\sqrt{\omega/\sigma_g}$. The fields are decreased with decreasing frequency ω or increasing ground conductivity σ_g. Similar comments apply to the transverse electric (*TE*) field components that may be derived from H_r. The *TE* fields will not be considered in more detail.

The component E_r^{hm} that is associated with transverse magnetic (*TM*) fields may be compared with the corresponding component E_r^{ve} of a vertical elec-

tric dipole using an expression of the E_r^{ve} fields derived under similar assumptions [Wait, 1957]. Thus

$$E_r^{hm}\Big|_n = -\frac{\sin\phi E_0^{hm}}{S_n E_0^{ve}} E_r^{ve}\Big|_n = \frac{ikI^{hm}da}{S_n I^{ve}ds}\sin\phi E_r^{ve}\Big|_n. \quad (40)$$

Provided that the fields of the two antennas are the same in the given direction over a perfectly conducting plane earth ($\sin\phi E_0^{hm}=E_0^{ve}$), (40) may be simplified to

$$\frac{E_r^{ve}}{E_r^{hm}}\Big|_n = -S_n. \quad (41)$$

Substituting the reciprocity relation (3) in (40) follows that $(E_r^{ve})/(\eta H_\phi^{ve})$ also satisfy (41). The latter ratio, obtained from (13), (15), and (16) of Wait [1960b], when used in conjunction with (3) provides an alternate way of obtaining (40) and subsequently (41).

A more exact treatment of the horizontal magnetic dipole fields cannot be expected to lessen the importance of *TM* relative to *TE* field components. A better approximation to the *TM* fields of the horizontal magnetic dipole may be worked out from the exact ratio E_r^{ve}/E_r^{hm} derived in the appendix. As long as the source and observation point are at equal radii r the ratio involves only the azimuthal θ functions explicitly. However, the accuracy of the roots S_n and of E_r^{ve} depends on the approximations to the azimuthal and radial functions.

Numerical values of S_n are available from mode calculations of the vertical electric dipole [Howe and Wait, 1957]. For $\sigma_g=\infty$ the values are as follows:

h/λ	3.5	3.5	7
$\pi\nu c/(\omega_p^2 h)$	0.1	0.01	0.1
1	$0.9975-i\,3\times10^{-4}$	$0.997-i\,10^{-3}$	$0.995-i\,6\times10^{-3}$
2	$.98-i\,3\times10^{-3}$	$.97-i\,8\times10^{-3}$	$.992-i\,4\times10^{-4}$
3	$.93-i\,0.01$	$.91-i\,9\times10^{-3}$	$.98-i\,10^{-3}$
4	$.87-i\,0.02$	$.85-i\,0.01$	$.985-i\,2\times10^{-3}$

The $n=0$ mode is severely attenuated for frequencies above a few kilocycles [Wait, 1957]. In the lower frequency range the first order perturbation solution of the modal equation may be used [Schumann, 1954], which gives

$$C_n=\frac{\omega_n}{\omega}+\frac{ic}{h\omega C_n}\left[\frac{1}{n_g}\sqrt{1-\frac{1-C_n^2}{n_g^2}}+\frac{1}{n_i}\sqrt{1-\frac{1-C_n^2}{n_i^2}}\right]. \quad (42)$$

The cutoff frequency of the nth mode is

$$\omega_n=\frac{n\pi c}{h}. \quad (43)$$

The perturbation method applies for $n\neq0$ if $C_n\approx\omega_n/\omega$. Substituting this for C_n^{-1} in the right-hand side of (42) and expanding the square root expressions

$$C_n\approx\frac{\frac{\omega}{\omega}+\frac{i_n c}{h\omega_n}\left[\left(\frac{1}{n_g}+\frac{1}{n_i}\right)-\frac{1}{2}\left(\frac{1}{n_g^3}+\frac{1}{n_i^3}\right)\right]}{1-\frac{ic}{2h\omega}\left(\frac{1}{n_g^3}+\frac{1}{n_i^3}\right)} \quad (44)$$

The latter expression may be also obtained by solving (42) as a quadratic and by subsequently ignoring the second order perturbation. Applying (14), (42) simplifies for $n=0$ and $|n_g|$, $|n_i|\gg1$ to

$$S_0=\sqrt{1+\frac{c}{hi\omega}\left[\frac{1}{n_g}+\frac{1}{n_i}\right]}. \quad (45)$$

For $n_g\gg n_i$ and $\sigma_i=1.2\times10^{-6}(\omega_p^2/\nu=1.35\times10^5)$ S_0 has the following values

ω	10^4	10^3	10^2
S_0	$1.04\,e^{-i2^\circ}$	$1.14\,e^{-i7^\circ}$	$1.46\,e^{-i6^\circ}$

The above tabulations show that $|S_n|$ or the ratio $|E_r^{ve}/E_r^{hm}|$ are decreased with increasing mode index n and, with the exception of S_0, also with decreasing frequency. The phase angle between the propagating modes of the two dipoles is negligible. The largest difference between the fields of the two dipoles may be expected for the lowest propagating frequencies or near the mode cutoff of $n\neq0$ modes. The values of S_n near the mode cutoff may be determined from an approximate solution of the modal equation. The first order perturbation solution is applicable only to the lower order modes. A different approximate solution is obtained by observing that S_n is small near the cutoff ($S_n=1/(2\,ka)$ at the mode cutoff for perfectly conducting ground and ionosphere according to Schumann [1954]). Hence one may look for solutions of the modal equation of the form

$$C_n=\sqrt{1-S_n^2}=1-\Delta_n \quad (46)$$

where $|\Delta_n|\ll1$. Substituting (46) in (12), letting $p=i$ and defining

$$L=\frac{\epsilon_0\omega}{\sigma_i}=\frac{\omega\nu}{\omega_p^2}, \quad (47)$$

results after neglecting terms with Δ_n^2 and higher powers of Δ_n in

$$R_i^n=i(2L-i-2\sqrt{L}\sqrt{L-i})\left(1-\frac{2\Delta_n\sqrt{L}}{\sqrt{L-i}}\right)=L_1+\Delta_n L_2. \quad (48)$$

Considering $n_g^2\gg1$, substituting (46) in (12) and letting $p=g$ gives

$$R_g^n\approx1-2\,n_g^{-1}\,e^{i\pi/4}(1+\Delta_n)=G_1+\Delta_n G_2. \quad (49)$$

The solution of the modal equation (21) becomes

$$\Delta_n=\frac{2i(kh-\pi n)-lnL_1G_1}{2ikh+\frac{L_2}{L_1}+\frac{G_2}{G_1}}. \tag{50}$$

It is simplified at the cutoff of the respective mode by applying (43) to

$$\Delta_n=-\frac{lnL_1G_1}{2in\pi+\frac{L_2}{L_1}+\frac{G_2}{G_1}}. \tag{51}$$

The calculated curves of $|S_n|$ and of its argument are depicted in figures 2 and 3 for infinite ground conductivity. $|S_n|$ is smallest for small values of $I=\pi c\underline{\nu}/(h\omega_p^2)$ (or for large ionospheric conductivities) and for the higher modes (higher cutoff frequencies $\omega_n=n\pi c/h$). The argument of S_n exhibits a relatively small variation with n. The increase of $|S_n|$ and of arg S_n with increasing I signifies the increase of mode attenuation with decreasing ionospheric conductivity. The increase of $|S_n|$ from the $\sigma_g=\infty$ to the $\sigma_g=0.07/h_{km}$ curve ($\sigma_g=10^{-3}$ mho/m for $h=70$ km) is less than 20 percent for the $n=1$ mode in figure 4. This relative increase is even smaller for the higher n values.

The amplitude ratio between electric and magnetic dipole fields in (41) will be larger for propagating modes than near the mode cutoff. The cutoff ratio is larger than 0.25 for the $n=1$ mode and larger than 0.15 for the first 8 modes with $I\gtrsim 0.01$ (or for $\sigma_i\geq 1.2\times 10^{-5}$ with $h=70$ km). It may be noted that this comparison is made for dipoles that exhibit equal far fields above a perfectly conducting plane ground in the absence of ionosphere. The knowledge of the above amplitude ratios would enter in a comparison of antenna radiation efficiencies. However, no such calculations have been attempted in this paper.

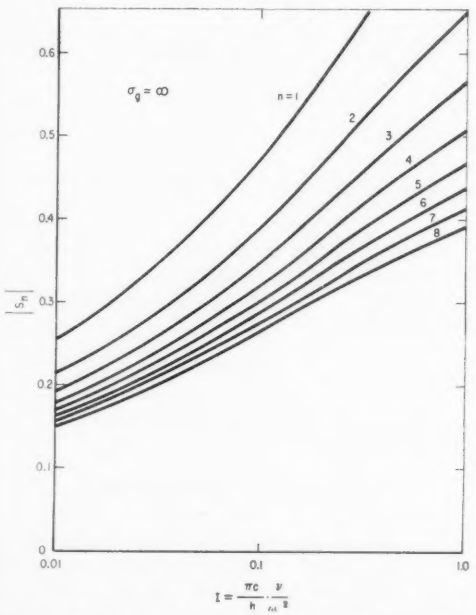


FIGURE 2. *Magnitude of the constant* S_n *near mode cutoff for perfectly conducting ground.*

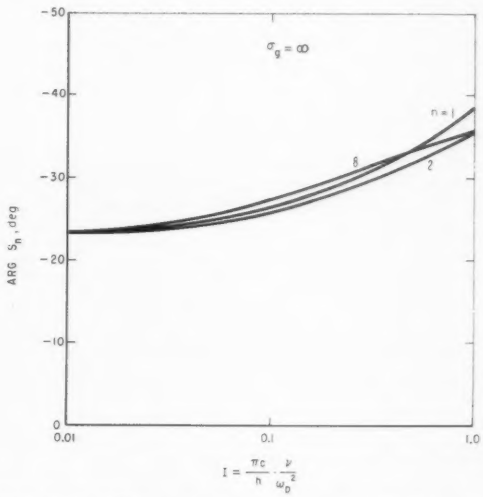


FIGURE 3. *Cutoff arguments of the constant* S_n *for perfectly conducting ground.*

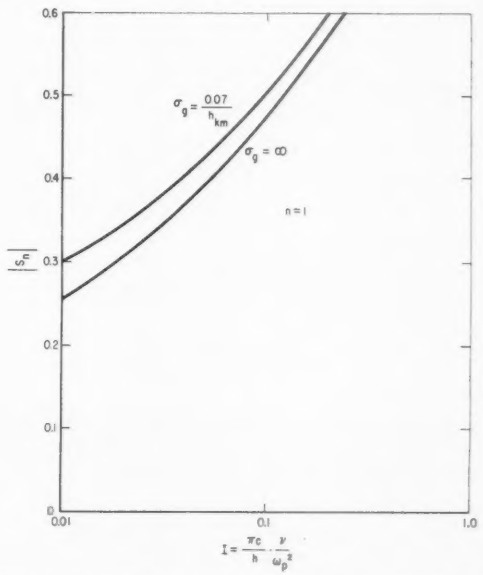


FIGURE 4. *Magnitude of the constant* S_n *near cutoff for the* n=1 *mode.*

5. Transient Fields

The transient fields of the horizontal magnetic dipole may be computed as the Fourier integral of (32) for a specified spectrum of the antenna current. For a step of current the far field approximation (32) gives

$$E_r^{hm}(\omega)=K\sqrt{i\omega}\sum_{n=0}^{\infty}\delta_n S_n^{0.5}\,e^{-i\frac{\omega d}{c}}S_n \tag{52}$$

with

$$K=\frac{\sin\phi\, I_0\, da\eta}{2hc\sqrt{2\pi ac\sin(d/a)}} \tag{53}$$

and with $\delta_0=0.5$ and $\delta_{n\neq 0}=1$.

The response of the $n=0$ mode may be reduced to tabulated Fourier integrals after approximating S_0 of (52) by a power series. With $n_g\gg n_i\gg 1$ (45) reduces to

$$S_0\approx\sqrt{1+\frac{1}{h\sqrt{i\omega\mu\sigma_i}}}. \tag{54}$$

Expanding $\sqrt{S_0}$ and S_0 in powers of $(1/\sqrt{i\omega})$ and ending the expansions with the quadratic terms results in

$$E_r^{hm}(\omega)=Ke^{d/(2c\beta^2)}\left[\sqrt{i\omega}+\frac{1}{2\beta}-\frac{3}{8\beta^2\sqrt{i\omega}}\right]e^{-\frac{i\omega d}{c}-\frac{\sqrt{i\omega}d}{c\beta}}, \tag{55}$$

where

$$\beta=2h\sqrt{\mu\sigma_i}. \tag{56}$$

Applying the transform pairs (806), (801), and (807) of Campbell and Foster [1948] gives

$$E_{ro}^{hm}(\tau)\approx\frac{Ke^{1/u-0.25/x}}{2\alpha^{1.5}\sqrt{\pi x}}\left[\frac{1}{2x^2}-\frac{1}{x}\left(1-\frac{1}{u}\right)-\frac{3}{u^2}\right] \tag{57}$$

where

$$x=\tau/\alpha, \tag{58}$$

$$\tau=(t-d/c)>0, \tag{59}$$

$$\alpha=\left(\frac{d}{\beta}\right)^2=\frac{\epsilon_0 d^2}{4\sigma_i h^2}=\frac{\nu d^2}{4\omega_p^2 h^2}, \tag{60}$$

$$u=\frac{2c\beta^2}{d}=\frac{8\sigma_i h^2}{d}\sqrt{\frac{\mu_0}{\epsilon_0}}. \tag{61}$$

The power series expansion of (54) is applicable to frequencies where $\omega>\omega_0=(h^2\mu\sigma_i)^{-1}$. Hence (57) will be most accurate for $\tau<<2\pi/\omega_0$ or for

$$x<<x_0=2\pi\left[\frac{2h^2\sigma_i}{d}\sqrt{\frac{\mu_0}{\epsilon_0}}\right]^2=\pi u^2/8. \tag{62}$$

Considering $h=90$ km, $d=10^4$ km, $\sigma_i=10^{-6}$ mho/m as an example, $x_0=2.3$ with $u=2.4$. The representation of the transient peak in figure 5 will be accurate even for this pessimistically low σ_i value at $d=10^4$ km. Higher σ_i and smaller d result in larger x_0 (and u), which improves the accuracy of the transient tail representations.

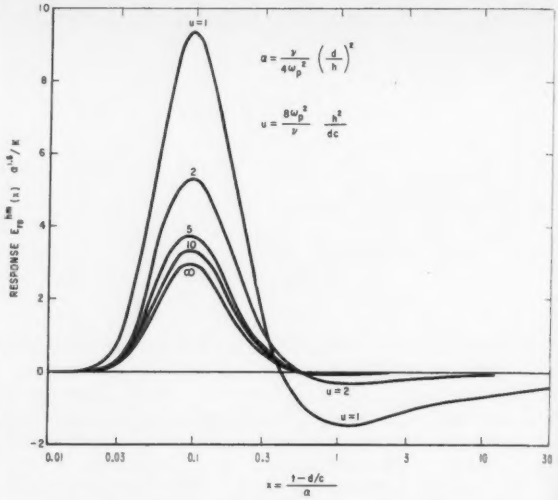


FIGURE 5. *Response to a current step of the* n=0 *mode of a horizontal magnetic dipole.*

It follows from reciprocity relation (3) that for equal current waveforms $I^{hm}(t)$ and $I^{ve}(t)$ in the *HM* and *VE* dipoles

$$E_r^{hm}(t)=-\mu_0 H_\phi^{ve}(t)\sin\phi da/ds. \tag{63}$$

A current in the *VE* dipole that is proportional to the time derivative of the current in the *HM* dipole

$$I^{ve}(t)=TI^{hm}(t), \tag{64}$$

results in

$$E_r^{hm}(t)=-\mu_0\sin\phi da H_\phi^{ve}(t)/(Tds). \tag{65}$$

The response $E_r^{hm}(t)$ to a current step $I_0u(t)$ should be proportional to the response $H_\phi^{ve}(t)$ to a current impulse $I_0T\delta(t)$. Assuming that (64) applies and assuming further that

$$S_0\approx 1 \tag{66}$$

(this is obviously incorrect at the lower frequencies), it follows from (40) that

$$E_r^{hm}(t)\approx-\sin\phi E_r^{ve}(t)da/(cTds). \tag{67}$$

The impulse response of a vertical electric dipole has been calculated considering factors proportional to finite powers of S_0 equal to unity [Schumann, 1952; Wait, 1960b]. The calculated impulse response of the *VE* dipole [Schumann, 1952; Wait, 1960b] will be compared with the step response of the *HM* dipole (57) by means of (65) and also by means of the approximate relation (67). The leading term

of the impulse response (42) of Wait [1960b] (eqs (38), (39), (40), and (42) of Wait [1960b] should be multiplied by $\pi/2$), when substituted in either (65) or (67) results in $E_r^{hm}(t)$ that is equal to (57) with the two terms proportional to u^{-1} and u^{-2} of the square brackets set equal to zero. There is no agreement between the higher response terms because of the S_0 approximation as discussed earlier. The impulse response of Schumann [1952], when substituted in (67) is the same as (57) with $u=\infty$.

The first order perturbation solution of the modal equation becomes inaccurate for the higher frequencies involved in calculating the transient response of the $n\neq 0$ modes. Even the simplest (and inaccurate) expressions of S_n that may be obtained from (44) and (14) for $|n_g|>>|n_i|>>1$ involve integrals similar to those encountered in the transient analysis of lossy rectangular waveguides [Cerillo, 1948]. The transient response of the $n\neq 0$ modes has not been calculated in this paper.

6. Appendix. Radial Electric Fields of Vertical Electric and Horizontal Magnetic Dipoles

The fields of a vertical dipole may be derived from a single scalar function. Applying (29) of Schumann [1954] to (3)

$$\frac{E_r^{ve}}{E_r^{hm}}=\frac{ds_1\,\frac{\partial}{\partial\theta}\left(\sin\theta\,\frac{\partial u_1}{\partial\theta}\right)}{rk^2\cos\phi\,da_2\sin\theta\,\frac{\partial u_2}{\partial\theta}} \tag{68}$$

where for u_2 the coordinates of the source and of the observation points are interchanged relative to u_1. As long as the source and the observation point are approximately at the same radius ($r_t\approx r_r$), both u functions are the same. The ratio of the vertical electric fields of horizontal magnetic and vertical electric dipoles for equal currents on a perfectly conducting plane ground is

$$\frac{E_0^{ve}}{E_0^{hm}}=\frac{ids_1}{kda_2}. \tag{69}$$

With

$$u_1\approx u_2\sim P_\nu\,(-\cos\theta) \tag{70}$$

(68) becomes

$$\frac{E_r^{ve}}{E_r^{hm}}=\frac{E_0^{ve}}{E_0^{hm}\cos\phi}\,\frac{i}{rk}\,\frac{(\nu+1)\sin\theta\,P_\nu\,(-\cos\theta)}{\cos\theta\,P_\nu\,(-\cos\theta)+P_{\nu-1}\,(-\cos\theta)}. \tag{71}$$

The earlier approximations will reduce (71) to (41). However, more accurate ν values [Wait, 1960a, ch. 12] and a better approximation to $P_\nu\,(-\cos\theta)$ may be used in (71) to obtain a more exact E^{ve}/E_r^{hm} ratio.

7. References

Bremmer, H., Terrestrial radio waves (Elsevier Publishing Co., New York, N.Y., 1949). See section III, 1 and 2.

Budden, K. G., The propagation of very low frequency radio waves to great distances, Phil. Mag. **44,** 504 (1953).

Campbell, G. A., and R. M. Foster, Fourier integrals for practical applications (D. Van Nostrand Co., Inc., New York, N.Y., 1948).

Cerillo, M., Transient phenomena in waveguides, M.I.T. Research Lab. of Electronics, TR No. 33 (Jan. 3, 1948).

Howe, H. H., and J. R. Wait, Mode calculations for VLF ionospheric propagation, Proc. Symposium on Propagation of VLF Radio Waves **3,** paper 36, Boulder, Colo. (1957).

Norton, K. A., The propagation of radio waves over the surface of the earth and in the upper atmosphere, pt. II, Proc. IRE **25,** 1203 (1937)

Schumann, W. O., Über die Ausbreitung sehr langer elektrischer Wellen und der Blitzentladung um die Erde, Z. Angew. Phys. **4,** 474 (1952).

Schumann, W. O., Über die Oberfelder bei der Ausbreitung langer, elektrischer Wellen im System Erde-Luft-Ionosphäre und 2 Anwendungen (horizontaler und senk-rechter Dipol), Z. Angew. Phys. **6,** 35 (1954).

Sommerfeld, A., Partial differential equations in physics (Academic Press, Inc., New York, N.Y., 1949).

Wait, J. R., The mode theory of VLF ionosphere propagation for finite ground conductivity, Proc. IRE **45,** 760 (1957).

Wait, J. R., Terrestrial propagation of VLF radio waves—a theoretical investigation, J. Research NBS **64D,** 153 (1960a).

Wait, J. R., Mode theory and the propagation of ELF radio waves, J. Research NBS **64D,** 387 (1960b).

Watson, G. N., The transmission of electric waves round the earth, Proc. Roy. Soc., Vol. **A95,** 546 (1919).

(Paper 65D3–133)

Publications of the National Bureau of Standards*

Selected Abstracts

Amplitude-probability distributions for atmospheric radio noise, W. Q. Crichlow, A. D. Spaulding, C. J. Roubique, and R. T. Disney, *NBS Mono. 23 (1960), 20 cents.*
Families of amplitude-probability distribution curves are presented in a form such that by using three statistical parameters of atmospheric radio noise, of the type published by the National Bureau of Standards, the corresponding amplitude-probability distribution may be readily chosen. Typical values of these parameters are given.

Low- and very low-radiofrequency model ionosphere reflection coefficients, J. R. Johler, L. C. Walters, and J. D. Harper, Jr., *NBS Tech. Note 69 (PB161570), (July 1, 1960) $2.00.*
The results of extensive computations performed during the course of a theoretical investigation of a sharply bounded model ionosphere for low- and very low-radiofrequency wave propagation are presented in the form of graphs and tables. *This Tech. Note supplements the work described in,* **On the theory of reflection of low- and very-low-radio frequency waves from the ionosphere,** *J. Research NBS* **64D,** *No. 3, 269 (May–June 1960).*

The relation of $h_{max}F2$ to M(3000)F2 and h_pF2, J. W. Wright and R. E. McDuffie, *J. Radio Research Laboratories* **7,** *No. 32, 409 (July 1960).*
Several easy methods of inferring directly from ionograms the height of the F region peak, using the quantities $M(3000)F2$ or h_pF2, are compared with values of $h_{max}F2$ obtained from $N(h)$ profiles. It is found that such methods are free from bias at night at low and medium latitudes, but that systematic differences are significant in the daytime, and at high latitudes at all times. Examples are given illustrating the incorrect diurnal variation of $h_{max}F2$ estimated from these simple parameters, in comparison with simultaneous values of h_{max} obtained from $N(h)$ profiles.

A test of a procedure for easy estimation of representative monthly electron density profiles for the ionosphere, J. W. Wright, *J. Geophys. Research* **65,** *No. 10, 3215–3217 (Oct. 1960).*
A recently proposed method for obtaining a representative monthly electron-density profile for a given hour of the day applies standard $N(h)$ procedures to a single representative virtual-height curve obtained in a special way from the individual daily virtual-height curves for that hour. The method is tested by comparing examples of the resulting profile with the average of the profiles for the individual days. The differences are small compared with the dispersion of the individual profiles about their mean. It is concluded that the method is valuable for obtaining a worldwide $N(h)$ morphology with a minimum expenditure of effort.

Spiral patterns in geophysics, V. Agy, *J. Atmospheric and Terrest. Phys.* **19,** *No. 2, 136–140 (Oct. 1960).*
Recent analyses of magnetic and ionospheric data, notably by A. P. Nikolski, have resulted in spiral "precipitation" patterns which lead the authors to claim support for Störmer's theory of the aurora. Although it may be possible to argue against these claims by attacking the methods of analysis and/or the Störmer theory itself, an entirely different approach is used in the paper presented here: an examination of pertinent points of Störmer's theory shows that the analytical spirals mentioned cannot be Störmer spirals and this conclusion holds regardless of the soundness of the analyses and, indeed, of the validity of Störmer's theory.

Supplementary world maps of F2 critical frequencies and maximum usable frequency factors, Donald H. Zacharisen, *NBS Tech. Note 2–2 (PB151361–2) (Oct. 1960) $3.50.*
This report supplements NBS Tech. Note 2 (April 1959), and completes the basic data required for F2-layer maximum usable frequency predictions. Prediction charts are given for the months of February, April, May, August, October, and November. Auxiliary charts are included to aid in predicting F2-layer MUFs.

The four parameters used for predicting MUFs are foF2 and the 4000 km MUF factor for a twelve-month running average Zurich sunspot number of 50, and the rates of change of foF2 and 4000 km MUF factor with sunspot number. The first three parameters are presented in map form for each even hour of Greenwich Mean Time. The fourth parameter is presented on a chart of geomagnetic latitude and local time.

The height of maximum luminosity in an auroral arc, F. E. Roach, J. G. Moore, E. C. Bruner, Jr., H. Cronin, and S. M. Silverman, *J. Geophys. Research* **65,** *No. 11, 3575–3580 (Nov. 1960).*
The height of maximum luminosity of an auroral arc is estimated from simultaneous observations at three stations in western United States during a night of general auroral activity (November 27–28, 1959). Photometrically this arc is characterized by a selective enhancement of the [OI] 6300 A line. From twenty-four individual measurements the height is found to be 412 km with a standard deviation of ± 23 km for one observation and ± 5 km for the mean. The geographical position of the arc, its orientation, and its movement during the night are discussed.

FM and SSB radiotelephone tests on a VHF ionospheric scatter link during multipath conditions, J. W. Koch, W. B. Harding, and R. J. Jansen, *IRE Trans. Commun. Systems* **CS–8,** *No. 3, 183–186 (Sept. 1960).*
Experiments have been carried out on an ionospheric-scatter link to observe the effects of long-delayed multipath signals, caused by F_2 propagated back scatter, on the intelligibility of voice communication. Frequency modulation and single-sideband modulation equipments were used for the tests. During periods when the back-scatter signal levels approached the level of the normal ionospheric-scatter signals, the frequency-modulation voice transmissions were unintelligible; however, under the same conditions, single-sideband voice communication intelligibility remained at almost 100 per cent although there was some loss in quality.

Radio refractometry, Jack W. Herbstreit, *NBS Tech. Note 66 (PB 161567) (July 1960) 50 cents.*
The optical refractive index is known to be determined principally by the temperature and pressure of the atmosphere, whereas the radio refractive index is, in addition, affected by the water content of the atmosphere, the relationship between these quantities being expressed in the following way:

$$N=(n-1)\ 10^6=\left(\frac{77.6}{T}\right)(P+[4810e/T]) \qquad (1)$$

where total air pressure P, and water vapor e, are in millibars, and the temperature T is in degrees Kelvin [Smith, 1953]. The quantities P, e, and T have long been routinely measured at the surface of the earth by standard weather bureau stations. For some time they have been measured from the surface up to great heights using balloon-borne radiosonde equipment at a large number of places over the earth's surface. More recently, equipment has been developed to measure rapidly and directly the radio refractive index of the atmosphere using radio techniques. The measurement of the radio refractive index properties of the atmosphere and the application to radio propagation problems is the subject of this paper.

Other NBS Publications

Journal of Research, Vol. 65A, No. 2, March–April 1961. 70 cents.

Mass spectra of some deuteroethanes. Edith I. Quinn and Fred L. Mohler.

Heats of hydrolysis and formation of potassium borohydride. Walter H. Johnson, Richard H. Schumm, Isa H. Wilson, and Edward J. Prosen.

Heat of combustion of borazine $B_3N_3H_6$. Marthada V. Kilday, Walter H. Johnson, and Edward J. Prosen.

Thermodynamic properties of thorium dioxide from 298 to 1,200 °K. Andrew C. Victor and Thomas B. Douglas.

Calculated energy dissipation distribution in air by fast ► electrons from a gun source. John E. Crew.

Vitrons as flow units in alkali silicate binary glasses. Leroy W. Tilton.

Tetragermanates of strontium, lead, and barium of formula type AB_4O_9. Carl R. Robbins and Ernest M. Levin.

Journal of Research, Vol. 65B, No. 1, January–March 1961. 75 cents.

On transient solutions of the "baffled piston" problem. F. Oberhettinger.

Special types of partitioned matrices. Emilie V. Haynsworth.

Bound for the *p*-condition number of matrices with positive roots. Philip J. Davis, Emilie V. Haynsworth, and Marvin Marcus.

Some computational problems involving integral matrices. Olga Taussky.

Computational problems concerning the Hilbert matrix. John Todd.

Index to the distributions of mathematical statistics. Frank A. Haight.

Selected bibliography of statistical literature, 1930 to 1957: IV. Markov chains and stochastic processes. Lola S. Deming and D. Gupta.

Building code requirements for reinforced masonry, NBS Handb. H74 (1960) 15 cents.

Amplitude, probability distributions for atmospheric radio noise, W. Q. Crichlow, A. D. Spaulding, C. J. Roubique, and R. T. Disney, NBS Mono. 23 (1960) 20 cents.

Standard frequencies and time signals from NBS stations WWV and WWVH, NBS Misc. Pub. 236 (1960) 10 cents.

Research highlights of the National Bureau of Standards, Annual Report, fiscal year 1960, NBS Misc. Pub. 237 (1960) 65 cents.

Quarterly radio noise data—December, January, February 1959–60, W. Q. Crichlow, R. D. Disney, and M. A. Jenkins, NBS TN18–5 (PB151377–5) (1960) $1.75.

Quarterly radio noise data—March, April, May 1960, W. Q. Crichlow, R. D. Disney, and M. A. Jenkins, NBS TN18–6 (PB151377–6) (1960) $1.75.

On the nature of the crystal field approximation, H. Goldberg and C. Herzfeld, NBS TN67 (PB161568) (1960) $2.50.

Vapor pressures of organic compounds in the range below one millimeter of mercury, E. E. Hughes and S. G. Lias, NBS TN70 (PB161571) (1960) 75 cents.

Some experiments on the deposition of gasses at 4.2 °K, T. Baurer, NBS TN73 (PB161574) (1960) $1.00.

Scattering of cobalt-60 gamma radiation in air ducts, C. Eisenhauer, NBS TN74 (PB161575) (1960) 75 cents.

Soviet research in field electron and ion emission, 1955–59; an annotated bibliography, T. W. Marton and R. Klein, NBS TN75 (PB161576) (1960) $1.25.

Oblique incidence receiving antenna array for a relative ionospheric opacity meter, A. C. Wilson, NBS TN78 (PB161579) (1960) 50 cents.

The dynamic compressibility of a rubber-sulfur vulcanizate and its relation to free volume, J. E. McKinney, H. V. Belcher, and R. S. Marvin, Trans. Soc. Rheology **IV,** 347 (1960).

Procedure for the determination of the noble metal content of dental gold alloys, H. J. Caul, W. S. Clabaugh and M. E. Susa, J. Am. Dental Assoc. **61,** No. 4, 339 (Oct. 1960).

Theory of anisotropic fluids, J. L. Ericksen, Trans. Soc. Rheology **IV,** 29 (1960).

Rapid frequency analysis of fading radio signals, J. M. Watts and K. Davies, J. Geophys. Research **65,** No. 8, 2295 (Aug. 1960).

On Stokes flow about a torus, W. H. Pell and L. E. Payne, Mathematika **7,** 78–92 (1960).

Reply to criticisms concerning Ireland contained in article, basic research in Europe, D. M. Gates, Science **128,** No. 3318, 1–9 (Aug. 1, 1958).

Thermal voltage converters for accurate voltage measurements to 30 megacycles per second, F. L. Hermach and E. S. Williams, Commun. Electron. **AIEE,** No. 49, 200 (July 1960).

Modern theories of materials, C. Truesdell, Trans. Soc. Rheology **IV,** 9 (1960).

Use of the incoherent scatter technique to obtain ionospheric temperatures, T. E. VanZandt and K. L. Bowles, J. Geophys. Research **65,** 2627 (Sept. 1960).

Highlights of the consultative committee on international radio (CCIR) activities in the field of radio propagation, J. W. Herbstreit, 9th Plenary Assembly of CCIR Proc. IRE **48,** No. 1, 45 (Jan. 1960).

Graphs for bivariate normal probabilities, M. Zelen and N. C. Severo, Ann. Math. Stat. **31,** No. 3, 619 (Sept. 1960).

Temperature stratification in a non-venting liquid helium dewar, L. E. Scott, R. F. Robbins, and B. W. Birmingham, Proc. 1959 Cryogenic Engr. Conf. **5,** (Sept. 2–4, 1959).

A microwave impedance meter capable of high accuracy, R. W. Beatty, IRE Trans. Microwave Theory and Tech. **MTT–8,** No. 4 (July 1960).

Variational treatment of electron-hydrogen atom elastic scattering, S. Geltman, Phys. Rev. **119,** No. 4, 1283–1290 (Aug. 1960).

Relaxation processes in multistate systems, K. E. Shuler, Phys. Fluids **2,** No. 4, 442 (July–Aug. 1959).

Comment on models of the ionosphere above h_{max} F2, J. W. Wright, J. Geophys. Research **65,** 2595 (Sept. 1960).

Some magnetoionic phenomena of the Artic E-region, J. W. Wright, J. Atmospheric and Terrest. Phys. **18,** 276 (Aug. 1960).

Tables of thermodynamic and transport properties of air, argon, carbon dioxide, carbon monoxide, hydrogen, nitrogen, oxygen and steam, J. Hilsenrath, C. W. Beckett, W. S. Benedict, L. Fano, H. J. Hoge, J. F. Masi, R. L. Nuttall, Y. S. Touloukian, and H. W. Woolley, 478 pages (Pergamon Press, Oxford, London, New York, and Paris, 1960).

Methods of nuclear orientation, E. Ambler, Progress in Cryogenics **2,** 235 (Heywood & Co., Ltd., London, England, 1960).

A conference on the propagation of ELF electromagnetic waves, J. R. Wait, Proc. IRE **48,** 1648 (Sept. 1960).

The strength of ten structural adhesives at temperatures down to minus 424 F., W. M. Frost, Proc. 1959 Cryogenic Engr. Conf. **5,** 375 (Sept. 2–4, 1959).

The Stokes flow about a spindle, W. H. Pell and L. E. Payne, Quart. Appl. Math. **18,** 257 (1960).

Possibility of detecting ionospheric drifts from the occurrence of spread F echoes at low latitudes, R. W. Knecht, Nature Letter **187,** 927 (Sept. 1960).

A compilation and correlation of the PVT data of normal hydrogen from saturated liquid to 80 degree K., R. B. Steward and V. J. Johnson, Proc. 1959 Cryogenic Engr. Conf. **5,** 548 (Sept. 2–4, 1959).

Pressure dependence of rotationally perturbed lines in the ultraviolet band spectrum of CN., H. P. Broida and S. Golden, Can. J. Chem. **38,** 1666 (1960).

Neighbor interactions and internal rotations in polymer molecules. IV. Solvent effect on internal rotarions, S. Lifson and I. Openheim, J. Chem. Phys. **33,** No. 1, 109 (July 1960).

Improvements in radio propagation prediction service, W. B. Chadwick, Elec. Engr. 1–4 (Sept. 1960).

Carbon resistance thermometry with mixed dc and rf currents,

J. J. Gniewek and R. J. Corruccini, Rev. Sci. Inst. **31,** No. 8, 899 (Aug. 1960).

Some SEAC computations of subsonic flows, P. Davis and P. Rabinowitz, book, Bergman's Linear Integral Operator Method in the Theory of Compressible Fluid Flow, by M. Z. Krzywoblocki, p. 148 (Wien, Springer-Verlag, Berlin, Germany, 1960).

Spatial distribution of energy dissipated by fallout beta rays, A. E. Boyd and E. E. Morris, Health Phys. **2,** 321 (1960).

Maser frequency stability, R. C. Mockler and J. A. Barnes, Proc. 13 Annual Frequency Control Symp., 583 (May 12–14, 1959).

Comment on a paper of Mori on time-correlation expressions for transport properties, M. S. Green, Phys. Rev. **119,** No. 3, 829 (Aug. 1960).

Acid-base equilibria in benzene at three temperatures. The comparative reactivities of a phenolic acid and a carboxylic acid with triethylamine with 1, 3-diphenylguanidine, M. M. Davis and M. Paabo, J. Am. Chem. Soc. **82,** 5081 (1960).

Absolute isotopic abundance ratio and the atomic weight of silver, W. R. Shields, D. N. Craig, and V. H. Dibeler, J. Am. Chem. Soc. **82,** 5033 (1960).

Flexural strength of specimens prepared from several uranium dioxide powders; its dependency on porosity and grain size and the influence of additions of titania, F. P. Knudsen, H. S. Parker, and M. D. Burdick, J. Am. Ceram. Soc. **43,** No. 12, 641 (Dec. 1960).

Green and purple sulfur: Electron-spin resonance studies, H. E. Radford and F. O. Rice, J. Chem. Phys. **33,** No. 3, 774 (Sept. 1960).

Correlation of an auroral arc with a subvisible monochromatic 6300 A arc with outerzone radiation on November 28, 1959, B. J. O'Brien, J. A. Van Allen, F. E. Roach, and C. W. Gartlein, J. Geophys. Research **65,** No. 9, 2759 (Sept. 1960).

Report on the standardization of pH and related terminology, R. G. Bates and E. A. Guggenheim, Intern. Union Pure and Appl. Chem. **1,** No. 1, 163 (1960).

Spectrum of ReF_3, J. C. Eisenstein, J. Chem. Phys. **33,** No. 5, 1530 (Nov. 1960).

Regulated power supply for instruments, W. V. Loebenstein, Electronics **33,** No. 48, 132 (Nov. 1960).

Effect of structure on the spectra emitted by solid nitrogen during electron bombardment, L. J. Schoen and H. P. Broida, J. Mol. Spectroscopy **5,** No. 5, 416 (Nov. 1960).

The compound $BaTiGe_2O_6$, C. R. Robbins, J. Am. Ceram. Soc. **43,** No. 11, 610 (Nov. 1960).

Chemical reactions of free radicals at low temperatures, R. A. Ruehrwein, J. S. Hashman and J. W. Edwards, J. Phys. Chem. **64,** 1317 (1960).

Experimental investigation of creep deflection of extruded and riveted I-beams, L. Mordfin and N. Halsey, NASA Tech. Note D–662 (Dec. 1960).

Spatial distribution of energy dissipated by fallout-rays, A. E. Boyd and E. E. Morris, Health Phys. **2,** 321 (1959).

Megaroentgen dosimetry employing photographic film without processing, W. L. McLaughlin, Radiation Research **13,** No. 4, 594 (Oct. 1960).

The reaction of hydrogen atoms with solid propene at low temperatures, R. Klein, M. D. Scheer and J. G. Waller, J. Phys. Chem. **64,** 1247 (1960).

The measurement of thermal conductivity, D. C. Ginnings, Book, Thermoelectric Materials and Devices, edited by I. B. Cadoff and E. Miller. Chapter 8, 113 (Reinhold Publ. Corp., New York, N.Y., 1960).

Programming for a closed-loop, manned-machine combined system, D. C. Friedman, Proc. Combined Analog-Digital Computer Systems Symp., December 16 and 17, 1960, Philadelphia, Pa., 12th Article (Dec. 1960).

Irrational power series, M. Newman, Proc. Am. Math. Soc. **11,** 699 (Oct. 1960).

Analytical study of creep deflection of structural beams, L. Mordfin, NASA Tech. Note D–661 (Dec. 1960).

Evaluation of ball bearing separator materials operating submerged in liquid nitrogen, W. A. Wilson, K. B. Martin, J. A. Brennan, and B. W. Birmingham, ASLE. ASME Lubrication Conf., October 17–19, 1960 (Boston, Mass.), Am. Soc. Lubrication Engrs. Preprint No. 60 LC–4 (1960).

Microwave spectrum, structure, and dipole moment of propane, D. R. Lide, Jr., J. Chem. Phys. **33,** No. 5, 1514 (Nov. 1960).

Topological derivation of the Mayer density series for the pressure of an imperfect gas, M. S. Green, J. Math. Phys. **1,** No. 5, 391 (Sept.–Oct. 1960).

Surface area determination of kaolinite using glycerol adsorption, K. H. Woodside and W. C. Ormsby, J. Am. Ceram. Soc. **43,** No. 12, 671 (Dec. 1960).

Magnetic resonance determination of the nuclear moment of tantalum–181 in $KTaO_3$, L. H. Bennett and J. I. Budnick, Phys. Rev. **120,** No. 5, 1812 (Dec. 1, 1960).

Aircraft storage batteries, W. J. Hamer, AIEE and Am. Inst. Elec. Engr. Trans. **79,** Pt. II, 1–11 (Sept. 1960).

Electrodeless passage of direct current through an electrolyte, A. Brenner, J. Electrochem. Soc. **107,** No. 12, 968 (Dec. 1960).

Preparation and thermal stability of tetrakis-(pentafluorophenyl)-silane and tris-(pentafluorophenyl)-phosphine, L. A. Wall, R. E. Donadio, and W. J. Pummer, J. Am. Chem. Soc. **82,** No. 18, 4846 (Sept. 1960).

Stress-strain relationships in yarns subjected to rapid impact loading. Part VI: Velocities of strain waves resulting from impact, J. C. Smith, J. M. Blanford, and H. F. Schiefer, Textile Research J. **30,** No. 10, 752 (Oct. 1960).

The centennial of spectrochemistry, W. F. Meggers, J. L. Tech, J. Opt. Soc. Am. **50,** No. 11, 1035 (Nov. 1960).

Rate of reaction of nitrogen atoms with ethylene, J. T. Herron, J. Chem. Phys. **33,** No. 4, 1273 (Oct. 1960).

Structure of the isobutane molecule; change of dipole moment on isotopic substitution, D. R. Lide, Jr., J. Chem. Phys. **33,** No. 5, 1519 (Nov. 1960).

Condensation coefficient of arsenic trioxide glass, A. B. Bestul and D. H. Blackburn, J. Chem. Phys. **33,** No. 4, 1274 (Oct. 1960).

Some mechanical properties of magnesium alloys at low temperatures, R. P. Reed, R. P. Mikesell and R. L. Greeson, Proc. 1959, Cryogenia Eng. Conf. **5,** 397 (Sept. 2–4, 1959).

Electron spin resonance studies of free radicals in irradiated materials, L. A. Wall, Symp. on Materials in Nuclear Application, Am. Soc. Testing Materials Spec. Tech. Publ. No. 276, p. 208 (1959).

**Publications for which a price is indicated (except for Technical Notes) are available only from the Superintendent of Documents, U.S. Government Printing Office, Washington 25, D.C. (foreign postage, one-fourth additional). The Technical News Bulletin and Basic Radio Propagation Predictions are available on a 1-, 2-, or 3-year subscription basis, although no reduction in rates can be made. Reprints from outside ournals and the NBS Journal of Research may often be obtained directly from the authors.*

Notice

The papers listed below will appear in early issues of RADIO PROPAGATION:

Almost fifty years of URSI. J. Howard Dellinger.

Power density requirements for airglow excitation by gyrowaves. V. A. Bailey.

On the validity of some approximations to the Appleton-Hartree formula. Kenneth Davies and G.A.M. King.

Amplitude and angular scintillations of the radio source Cygnus-A observed at Boulder, Colorado. R. S. Lawrence, J. L. Jespersen, and R. C. Lamb.

Digital methods for the extraction of phase and amplitude information from a modulated signal. R. S. Lawrence, J. L. Jespersen, and R. C. Lamb.

Comparison between mode theory and ray theory of VLF propagation. H. Volland.

Antenna coupling error in direction finders. Charles W. Harrison, Jr.

The electrically short antenna as a probe for measuring free electron densities and collision frequencies in an ionized region. R. W. P. King, C. W. Harrison, Jr., and D. H. Denton, Jr.

Effect of multiple atmospheric inversions on tropospheric radio propagation. F. H. Northover.

A few observations of the perturbations in the phase of the low-frequency ground wave. J. M. Ross and J. E. Kirch.

Smooth earth diffraction calculations for horizontal polarization. L. E. Vogler.

On the theory of mixed-path ground-wave propagation on a spherical earth. James R. Wait.

Reflection from a sharply bounded ionosphere for VLF propagation perpendicular to the magnetic meridian. Douglas D. Crombie.

Resonance of the space between earth and ionosphere. H. Poeverlein.

Experimental study of inverted L-, T-, and related transmission-line antennas sheila Prasad and Ronold W. P. King.

A note concerning the excitation of ELF electromagnetic waves. James R. Wait.

UNITED STATES
GOVERNMENT PRINTING OFFICE
DIVISION OF PUBLIC DOCUMENTS
WASHINGTON 25, D.C.

OFFICIAL BUSINESS

PENALTY FOR PRIVATE USE TO AVOID
PAYMENT OF POSTAGE, $300
(GPO)

JOURNAL OF RESEARCH of the National Bureau of Standards Vol. 65D, No. 3 May–June 1961
D. Radio Propagation

Contents

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington 25, D.C. Price 70 cents (single copy)
Subscription price: $4.00 a year; $0.75 additional for foreign mailing.
UNITED STATES GOVERNMENT PRINTING OFFICE, WASHINGTON : 1961